重力透鏡現象的資料探勘

陳詩湧、葉永烜

國立中央大學天文研究所

摘要

　　依據廣義相對論，光線的行進路徑會受重力影響而彎曲。若遙遠的星體與地球之間有大質量的物質存在，在類似光線經過透鏡聚焦的情形下，我們會觀測到圓弧狀的天體或是雙重甚至多重影像。這種放大現象可以幫助我們更了解高紅移且微弱的星系。

　　目前大部份的天文學家用目視或基於星系團的稠密度來搜尋重力透鏡事件，而我們則利用星體形狀資訊進行物體的辨識。由於這種新的思維方法是基於形狀參數，理論上是可以偵測出由暗星系團或暗物質所造成的重力透鏡現象。

　　我們根據圖像的『第二階中心矩』守恆作出『橢圓形誤差』的形狀參數之定義，並使用它來篩選星體。接下來是要對可疑的星體做線性和弧形回歸。最後根據以上兩種回歸的誤差來判別它是否是重力透鏡作用的現象。由於我們所使用的回歸法和形狀參數，與圖像的矩函數有關係，在資料有效的運用下使得整個運算過程很快。

Data Mining for Gravitational Lenses

Tan Sze-Yeong, Ip Wing-Huen

Institute of Astronomy, National Central University

Abstract

　　The discovery of gravitational lensing events are very import in the study of cosmology. Unlike the traditional method of discovering the strong lenses which based on direct visual inspection on over-dense region of galaxy clusters or photometry of caustic sources, our method is based on morphology. Since our method is shape-motivated, it will not only relief the task of astronomers but also capable of detecting the "dark lenses".

　　We have defined a shape parameter to select the candidates efficiently. Through error analysis the arcs produced by lenses can be identified .The algorithm developed in this work has been tested on two of the gravitational lensing events found in the RCS and proved to be successful.

關鍵字（Key words）：重力透鏡（gravitational lensing）、矩分析（moments analysis）、merging galaxies、arc fitting

1. Introduction

1.1 Gravitational Lensing as Cosmological Probe

The number of gravitational lenses we expect to see depends very muchon the large-scale geometry and the evolutionary history of the universe. Thus the lens statistics can be used to constrain the cosmological parameters by comparing the number of lenses to that predicted by various models.
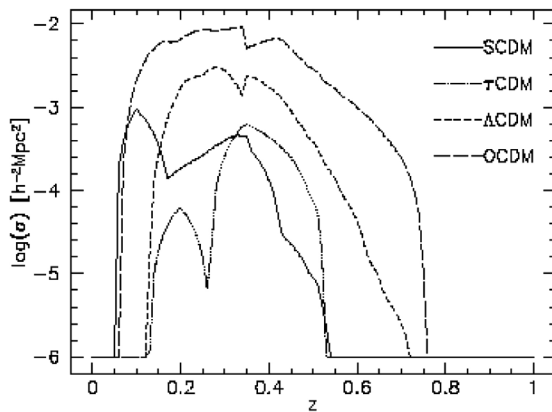

Fig 1. The maxima of the cross sections for four different cosmologies differ by more than an order of magnitude. The Figure shows that τCDM model produce the fewest arcs, and clusters in the open model produce the most. The redshift ranges where $\sigma(z)>0$ are larger in O/ΛCDM than in S/τCDM. (Bartelmann et al. 1998).

Bartelmann et al (1998) showed that the influence of the evolutionary history of cosmology on arc statistics is dramatic: for large arcs with length-to-width ratio ≥ 10, the optical depth is largest for OCDM($\Omega_0=0.3$, $\Omega_\Lambda=0$). The optical depth is lower by about an order of magnitude for ΛCDM, and by about two orders of magnitude for s/τCDM, relative to OCDM.

The average cross sections evaluated by averaging the cross section over the three projection directions for different cosmologies are shown in Figure 1.

2. Methodology

2.1 Basic Idea - Recognition by Fitting Error

Astronomers has been urged to do more lens surveys since the number of discovered lenses is still too few. However, searching the gravitational lensing events in large amount of wide field images is a laborious and time-consuming task. Therefore, the development of fast and reliable algorithm is more desirable. Since galaxy clusters are the most effective lenses for the occurrence of lensing events, searching for the over-dense region becomes a good strategy (Gladder and Yee, 2002 & Gladder et al. 2003). Alternatively, we can approach this problem by means of shape analysis.

We attempt to recognize the shape with an bimodal approach; a line and an arc are fitted to the unclassified object. Since fitting always comes with errors, we could just compare the errors and find out which kind of fitting would lead to a smaller amount of error; it means that the object is best described by that method. It could be shown that the principal axis along the semi-major axis of the objects is the best fitted line. The fitting error is thus the principal moment of inertia along semi-minor axis, a common parameter of the moment analysis.

Our algorithm involves image thresholding, moment (shape) analysis and arc fitting. There are described in the follow sections.

2.2 Image Segmentation by Local Thresholding

In local thresholding, the original image is

partitioned into many smaller sub-images and a threshold (*sky backgroud*+1.5σ_{sky}) is applied to each of these sub-images. Furthermore, the effect of discontinuity in graylevel is reduced by a statistical approach. For every sub-images, five threshold values are obtained by re- computing the background intensity for its four adjacent regions and an additionally enlarged region in situ. The final value is simply taken as the median of those five threshold values.

2.3 Moment Analysis and Shape Parameters

2.3.1 Definition of Moments

For an image *f(x,y)*, the moment of order *(p+q)* is defined as

$$m_{pq} = \sum \sum x^p y^q f(x, y)$$

In order to make the moments independent of the image reference system, they are often evaluated with respect to the first moment (*centroid*). These moments are called *central moments* and are defined as

$$\mu_{pq} = \sum \sum (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (1)$$
$$p, q = 0,1,2,3,...$$
$$\bar{x} = \frac{m_{10}}{m_{00}}; \quad \bar{y} = \frac{m_{01}}{m_{00}}$$

One can easily verify that $\mu = m_{00}$ and $\mu_{10} = \mu_{01} = 0$.

One of the simplest and widely used shape parameter defined by moments is called principle moments of inertia. They are a measure of variance of image intensity distribution about the centroid under a new reference frame such that μ_{11} is vanished. The mathematical relations of μ_{20}, μ_{02} and μ_{11} to the principle moments of inertia (I_1, I_2) are given below:

$$I_1 = \frac{\mu_{20} + \mu_{02} + \left[(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2\right]^{1/2}}{2}$$

$$I_2 = \frac{\mu_{20} + \mu_{02} - \left[(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2\right]^{1/2}}{2}$$

2.3.2 Best-Fit Ellipse

Best-fit Ellipse is an ellipse whose second geometrical moment equals to that of the object. According to the moments of a general ellipse,

$$I_{min} = \frac{\pi}{4} ab^3 \quad (2)$$
$$I_{max} = \frac{\pi}{4} a^3 b$$

where a and b represent the lengths of the semi-major and semi-minor axes of a best fitted ellipse, respectively. They can be obtained readily with given I_{min} and I_{max}.

$$a = \left(\frac{4}{\pi}\right)^{1/4} \left[\frac{(I_{max})^3}{I_{min}}\right]^{1/8} \quad (3)$$

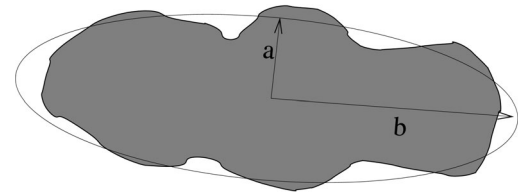$$b = \left(\frac{4}{\pi}\right)^{1/4} \left[\frac{(I_{min})^3}{I_{max}}\right]^{1/8}$$

Fig 2. An arbitrary shape super-imposed with its best-fit ellipse.

2.3.3 Dissimilarity

We define the dissimilarity of an object to an ellipse as below

$$dissimilarity = \frac{area(O \cup E) - area(O \cap E)}{area(O \cup E)} \quad (4)$$

O denotes the area enclosed by the object and *E* is for its best fitted ellipse.

This parameter is always bounded by 0 & 1.

Equation 4 could be expressed as

$$dissimilarity = 1 - \frac{area(O \cap E)}{area(O \cup E)}$$

For an object which does not intersect with its best-fit ellipse, it has dissimilarity of 1. While the object is a *perfect ellipse*, $area(O \cup E) = area(O \cap E)$; *it has dissimilarity of* 0.

It is worthwhile to mention that error becomes larger when the best-fit ellipse meets an curved arc-let. This is because the centroid of curved object tends *to shift outwardly*. As shown in Figure 3, the reconstructed ellipse is shifted to one side leaving a considerable amount of non-overlapping area.
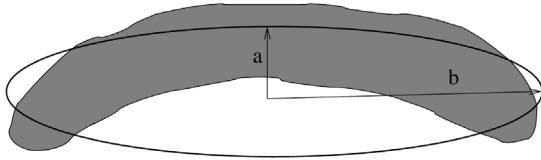

Fig 3. An arc super-imposed with its best-fit ellipse.

2.3.4 Examples

We have constructed a pair of ellipses with different ellipticities and a circular arc as test shapes for our shape parameter. They were corrupted with noises in different levels before performing the shape analysis. The results are shown in Figure 4. Best-fit ellipses were rendered in light gray and stacked under the test objects for demonstration.

2.4 Selection of Candidates by Using Dissimilarity and Eccentricity

A typical scatter-plot of a RCS image is depicted in Figure 5. From the diagram we could see that most of the stellar objects have a fairly low dissimilarity (~.0.06) and eccentricity around 0.2; it means that most of
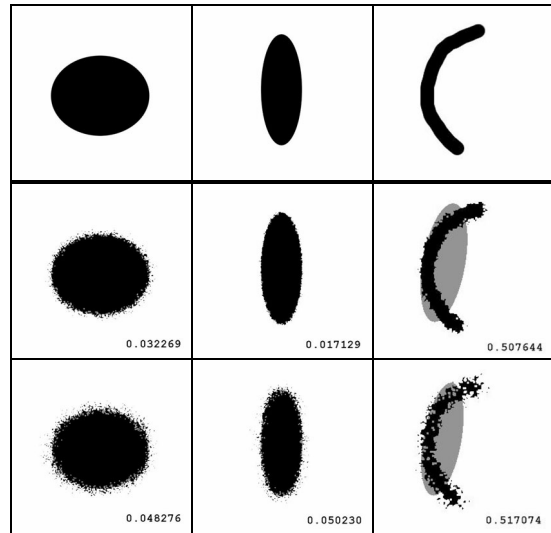

Fig 4. The figure shows two test objects. Noises with different extent was added to the test objects and best-fitted ellipse for each case were rendered in light Gray. The measurement of dissimilarity are also stated.

the objects look like regular ellipse. Note that the eccentricity we use here is $1 - a/b$ where *a* and *b* are the length of semi-major and semi-minor of the best fit ellipse. The density of the data points falls with the increase of dissimilarity and eccentricity. By experiences, we consider the object as a candidate if its dissimilarity > *0.15* or eccentricity > *0.65*. Under this rule, about 13% of the objects will be extracted as candidates for further analysis.
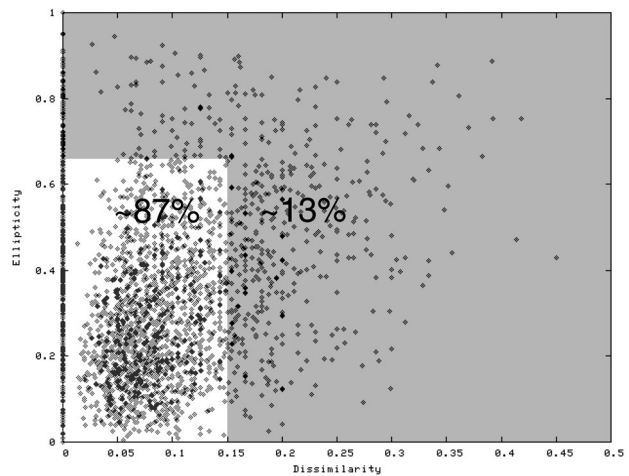

Fig 5. The typical scatter-plot of a RCS image in which the x-axis is dissimilarity and the y-axis is eccentricity.

2.5 Estimation of the Circular Arc Center and Its Radius

2.5.1 Landau's Algorithm

In this method, *the sum of the squares of the geometrical distances between data points and the center of a circle are optimized in least squares sense*, this is so the called *geometric fit*.

Given a set of points $(x_1, y_1), \cdots, (x_N, y_N)$, and $u_1, \cdots, u_N$ denoted as vectors to each of the points and a circle with a radius of R and a center at $r = (r_x, r_y)$; the error $\overrightarrow{\Delta r_i}$ can be defined by

$$\overrightarrow{\Delta r_i} = \vec{u}_i - \vec{r} - R \frac{\vec{u}_i - \vec{r}}{\left| \vec{u}_i - \vec{r} \right|} \qquad (5)$$

The mean square error is thus

$$e\left(R, \vec{r}\right) = \frac{1}{N} \sum_i \left(\Delta r_i\right)^2 = \frac{1}{N} \sum \left(\vec{u}_i - \vec{r} - R \frac{\vec{u}_i - \vec{r}}{\left| \vec{u}_i - \vec{r} \right|} \right)^2$$
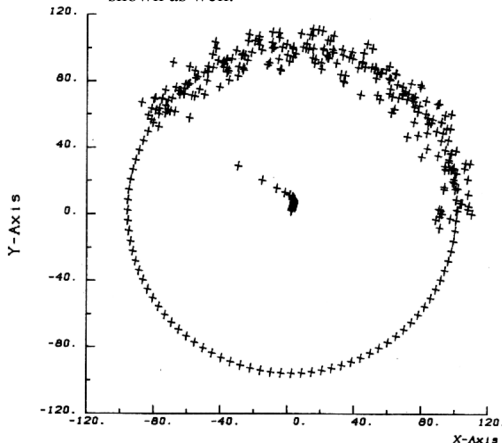
After minimizing the function $e\left(R, \vec{r}\right)$ with respect to R and $\vec{r}$, we get

$$R = \frac{1}{N} \sum_i \left| \vec{u}_i - \vec{r} \right| \qquad (6)$$

$$\vec{r}_c = \frac{1}{N} \sum_i \left(\vec{u}_i - R \frac{\vec{u}_i - \vec{r}}{\left| \vec{u}_i - \vec{r} \right|} \right) \qquad (7)$$

Since this is a *nonlinear* least squares problem, the solution will be solved by iteration.

Fig 6. This figure demonstrate the fitting of circular arc with Landau's algorithm, the location of the center at each iteration is shown as well.

2.5.2 Thomas-Chan's Algorithm

The algebraic solutions for the circular fitting are exact but not accurate. They are used as the initial values for the iterative Landau's algorithm (geometric fit).

The algebraic fit we adopted in this study is the Thomas-Chan's Algorithm. Instead of minimizing the sum of the squares of geometric distances, this method minimizes the sum of the squares of algebraic distances

$$e(R, \bar{x}, \bar{y}) = \sum_{i=1}^{N} \left[R^2 - \left\{ (x_i - \bar{x})^2 + (y_i - \bar{y})^2 \right\} \right]^2 \qquad (8)$$

3. Data Processing

3.1 The Red-Sequence Survey (RCS)

The Red-Sequence Cluster Survey(RCS) is a galaxy cluster survey designed to provide a large sample of optically selected $0.1 < z < 1.4$ clusters, spread over a wide range of RA and DEC in order to facilitate follow-up. The images was acquired at the CFHT and CTIO 4m telescopes using mosaic camera. It covered 100 square degrees of two colors (R and z') imaging, with a 5σ depth $\sim$ 2mag past M^* at $z = 1$. The primary scientific drivers of this survey are a derivation of Ω_m and σ_8 and a study of cluster galaxy evolution with a complete sample.

3.2 Incidence of Strong-Lensing in the RCS Survey

There are a total of eight strong-lensing systems found in RCS datasets (Gladders et al. 2003). The arcs were detected by visual examination of the over-dense regions with red galaxies. Two of the strong-lenses are discussed

here, they will be used as test images for our system.

3.3 RCS 0224.5-0002

This is the most obvious case of strong lensing in the RCS. At a redshift as high as $z = 0.773$, it is the highest redshift cluster yet known with significant strong lensing. The outermost large arc in RCS 0224.5-0002 has been spectroscopically confirmed to be at $z = 4.8786$.

The RCS 0224.5-0002 is shown to have a central velocity dispersion of ~ 1000 kms^{-1} by using singular isothermal sphere model. This estimation means that it is one of the most massive galaxy clusters known at $z \sim 0.8$.
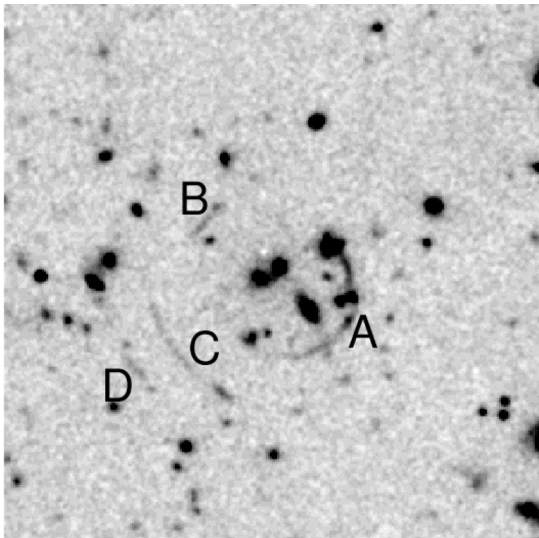

Fig 7. RCS 0224.5-0002, the strongest lensing in the RCS. The four putative lensed sources are label "A" - "D".

3.4 RCS 1419.2+5326

RCS 1419.2+5326 is a spectroscopically confirmed $z \sim 0.64$ cluster which produces two obvious giant arcs. However, the arcs are very different in radial position and the fainter one is marginally redder than the other. As a result, it has been suggested that there are correspond to

two different background sources, with the fainter one pertaining to a source at higher redshift than the brighter one.
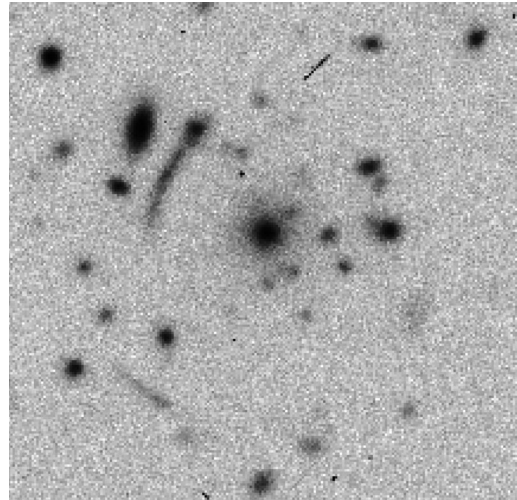

Fig 8. RCS 1419.2+5326, a strong-lensing system with spectroscopically confirmed z=0.64 cluster and two obvious arcs.

4. Data Processing Pipeline

The input images are converted into binary form after local thresholding as described in section 2.2 and followed by an morphology-opening operation with a 3×3 square block structure element. Geometrical moments and shape parameters are determined for objects larger than 30 pixels. They will be designated as candidates if their dissimilarity are greater than 0.15 and eccentricity higher than 0.65 (Figure 5).
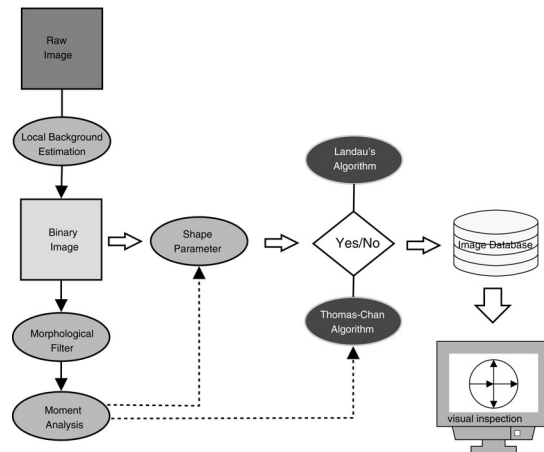

Fig 9. Data processing pipeline

- 15 -

The candidates are then further analyzed by arc-fitting algorithm. *Those with their arc-fitting error greater than their own principle moment of inertia are classified as curved objects.* Finally, a simple elimination rules is applied to lower the number of spurious detections before storing in image database. Those rules are:

(1) Discard object if its *effective width* is smaller than 3.5 pixels.

(2) Discard object if its fitted radius is smaller than 17 pixels.

(3) Discard object if its radius falls into the object itself. This could remove artifacts caused by bright stars.

Finally, the candidates are verified visually by retrieving the data from the image database.

5. Results of the Test Images

5.1 Result of RCS 0224-0002

Our program detected 177 candidates in RCS 0224-0002 and 5 of them were suggested to be possible arcs.

5.2 Result of RCS 1419.2+5326

For RCS 1419.2+5326, 132 candidates has been selected and 10 of them were suggested to be possible arcs. As shown in Figure 11, the big arc was recognized. However, we failed to detect the smaller arc as it is morphologically
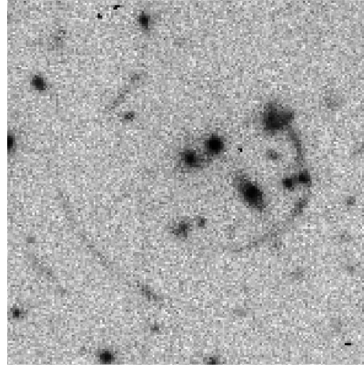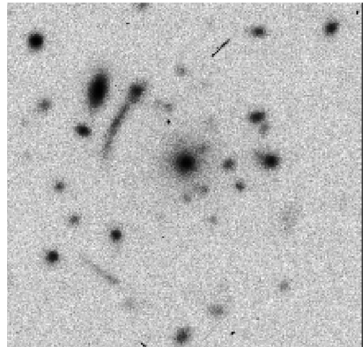

Fig 10. Detection of the primary arc in RCS 0224-0002.

Fig 11. Detection of the giant arc in RCS 1419.2+5326.

Fig 12. the faint and curved object in the image of RCS 1419.2+5326.

inconspicuous. Incidentally, one faint and curved object without cluster core was detected in this image also.

5.3 Parameter Spaces of The Detected Arcs

The typical scatter-plot of a RCS image imposed with the detected arcs mentioned before is depicted in Figure 13. It shows that the arcs have high dissimilarity and ellipticity compare to the ordinary stellar objects. Both of them have

Fig 13. The typical scatter-plot of a RCS image imposed with the detected arcs.
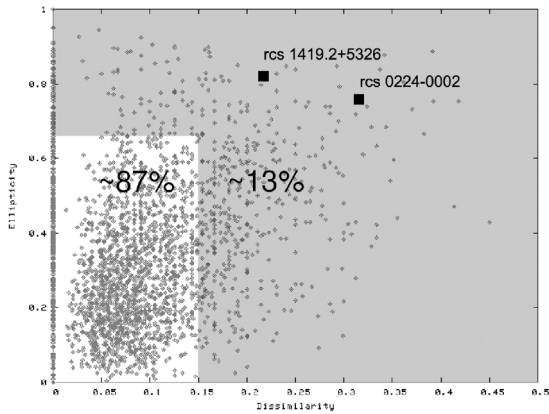
been captured successfully by our selection rules on parameter spaces.

6. Discussion

Our program was written in ANSI C and compiled with GNU C/C++ Compiler version 3.3.2. The processing time for a single image is about 13 s on a Linux platform with Pentium-4 2.4GHz and 1GB of memory.

The algorithm is fast because of the following reasons:

(1) The arc fitting algorithm is only performed on suspicious candidates which are only a fraction of the total number of the detected objects.

(2) The determination of the best-fit ellipse and dissimilarity are straight forward and require very low computation power.

(3) The convergence rate of the Landau's algorithm is improved by taking the algebraic solution as the initial values.

In spite of the fast processing speed of our program, the number of possible arcs suggested by the system is still too high. The false alarms are mainly due to the blended objects (close encounter). Therefore a deblending procedure is indispensable.

Due to the selection of threshold value, in some cases (Figure 10) some arcs does not in the corresponding binary image. One possible ways to improve this problem are:

(1) to develop a new algorithm to determine the optimal threshold, such that lesser amount of faint structures will be missed.

(2) to improve the accuracy of background estimation by some other schemes.

Reference

Bartelmann, M., Huss, A., Colberg, J. M., Jenkins, A., and Pearce, F., 1998, *A&A*, 330, 1.

Davies, D. H., and Dance, D. R., 1990, *Phys. MED. Biol.*, 35, 1111.

Gladders M., Henk, H, Yee, H.,. Hall,, P and Barrientos, L., 2003, *ApJ*, 593, 48.

Gladders M., Yee, H., and Ellingson, E., 2002, *ApJ*, 123, 1.

Gladders M., and Yee, H., 2000, *ApJ*, 120, 2148.

Henk H., Yee, H.,. and Gladders, M., 2002, *ApJ*, 577, 595.

Landau U., 1987, *Computer Vision, Graphics and Image Processing,* 38, 317.

Thomas S. M., Y. T. Chan, 1989, *Computer Vision, Graphics and Image Processing*, 45, 362.