A Spatial-Color Layout Feature for Representing Galaxy Images

Yin Cui[†], Yongzhou Xiang[†], Kun Rong[†], Rogerio Feris[§], Liangliang Cao[§]
[†]Department of Electrical Engineering, Columbia University
[§]IBM T. J. Watson Research Center
{yc2776, yx2211, kr2496}@columbia.edu, {rsferis, liangliang.cao}@us.ibm.com

Abstract

We propose a spatial-color layout feature specially designed for galaxy images. Inspired by findings on galaxy formation and evolution from Astronomy, the proposed feature captures both global and local morphological information of galaxies. In addition, our feature is scale and rotation invariant. By developing a hashing-based approach with the proposed feature, we implemented an efficient galaxy image retrieval system on a dataset with more than 280 thousand galaxy images from the Sloan Digital Sky Survey project. Given a query image, the proposed system can rank-order all galaxies from the dataset according to relevance in only 35 milliseconds on a single PC. To the best of our knowledge, this is one of the first works on galaxy-specific feature design and large-scale galaxy image retrieval. We evaluated the performance of the proposed feature and the galaxy image retrieval system using web user annotations, showing that the proposed feature outperforms other classic features, including HOG, Gist, LBP, and Color-histograms. The success of our retrieval system demonstrates the advantages of leveraging computer vision techniques in Astronomy problems.

1. Introduction

In the years of 2003-2004, the Hubble Space Telescope took a single 12-day exposure of $1/13,000,000$ of the entire sky. With around $10,000$ galaxies, the resulting "Hubble Ultra-Deep Field" suggests that the entire universe may contain up to 10^{11} galaxies [29]. Aiming at exploring the tremendous number of galaxies in the universe, the Sloan Digital Sky Survey (SDSS) project [18] has already collected almost 1 million images of galaxies to date. It is expected that the SDSS will collect more than 50 million galaxy images in the near future [7].

Manually processing and categorizing such a tremendous number of galaxy images would not be feasible. Thus, a new field called Astroinformatics is emerging at the intersection of Astronomy and Computer Science to help processing and mining information from huge astronomical data archives. In this context, this paper addresses the problem of large-scale galaxy image retrieval. Given an input galaxy image, our goal is to efficiently find similar galaxies in a given database containing a large number of galaxies (see Figure 1). As we will detail later in Section 6, designing such a system can be useful for astronomers in several applications.

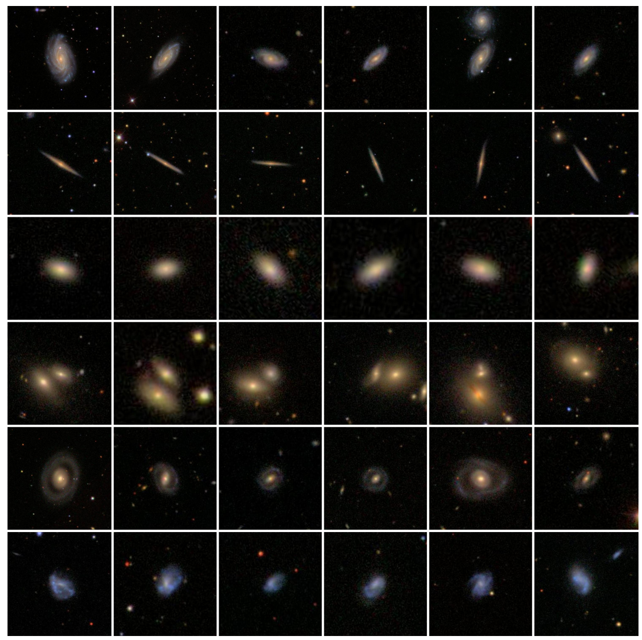
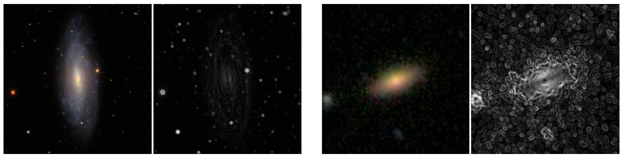


Figure 1. Galaxy image retrieval results using our method on SDSS dataset. From top to bottom, we give 6 different types of galaxies, which are Spiral, Edge-on, Smooth (Elliptical), Merging, Ring and Irregular corresponding to each of 6 rows respectively. Within each row, for a given query in the leftmost column, the top 5 retrieved galaxies are shown on the right-hand side.

Astronomical images are usually very noisy and contain a large area of dark background, posing challenges to automated methods that aim to measure galaxy similarity. Commonly used features in Computer Vision (*e.g.* LBP [20], Gist [21], Color-histogram [24], HOG [4]) are usually de-



(a) Spiral galaxy image and corresponding gradient magnitude (b) Elliptical galaxy image and corresponding gradient magnitude

Figure 2. Shape and gradient are not enough to capture representative information for spiral and elliptical galaxies.

signed for general natural images and cannot handle noisy and blurred galaxy images well. For example, Figure 2 shows two spiral and elliptical galaxies, respectively, from the SDSS dataset and their corresponding gradient magnitudes computed by the Difference of Gaussians (DoG) filter. Note that the gradient is noisy and fails to capture the galaxy global and local structures.

In order to capture the unique characteristics of galaxy images, this paper develops a new feature representation. Basically, we first design a detector that can segment the galaxy and remove the background. Then we use a spatial-color layout feature descriptor to capture localized information about color and shape of galaxies. Since current Astronomy studies show that local color and shape of a galaxy are correlated in terms of galaxy formation and evolution [8], localized color distributions could be a very useful cue in designing feature descriptors for galaxy images.

By combining our new feature with a state-of-the-art hashing technique [14], we implemented a system for large-scale galaxy image retrieval. Given a query galaxy image, our system can rank all galaxy images from a dataset with $283,971$ images by their similarities to the query in only 35 milliseconds using MATLAB running on a Intel 3.2 GHz Quad-core PC. A reliable evaluation system has been built. The evaluation results validate the success of our proposed feature over classic general-purpose features.

The main contributions of our paper lie in three aspects. Firstly, we proposed **a new feature** (detector and descriptor) for galaxy images that outperforms commonly used classic features in Computer Vision; secondly, using the proposed feature, we implemented **an efficient large-scale galaxy image retrieval system**. We are not aware of a similar system in the literature; thirdly, we proposed **a method to evaluate the performance of a galaxy retrieval system** using annotations collected from the internet.

The paper is organized as follows: Section 2 describes the related work for galaxy image feature and retrieval, from the perspective of both Astronomy and Computer Vision. In Section 3, we describe the proposed spatial-color layout feature detector and descriptor in details. The method we used for image retrieval is discussed in Section 4. In Section 5, we first introduce the evaluation system using crowd-sourced user annotations. Then we compare experimental results of our feature and other classic features. The analysis of our feature is also introduced in this section. Section 6 covers other applications of our work. Section 7 concludes the paper and suggests some directions for future research.

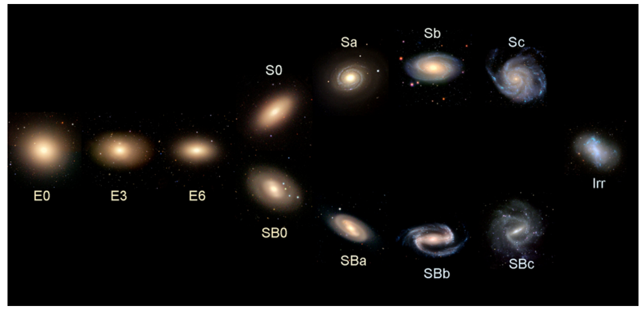


Figure 3. The Hubble Sequence.

2. Related Work

In this section we will review related work in galaxy image representation and retrieval. To the best of our knowledge, there is only a little work done on galaxy image retrieval using computer vision techniques and no prior work has been done with the same purposes as ours.

In Astronomy, the most widely used galaxy representation is known as the Hubble Sequence, a morphological classification scheme for galaxies proposed by Edwin Hubble in 1926 [11, 12]. Hubble groups regular galaxies into 3 broad classes in his classification system: elliptical, spiral and irregular. In each class, galaxies are further divided into several subclasses based on their shapes (Figure 3) (image credit: Department of Physics, University of Oregon). Up to now, the Hubble Sequence has become a common standard in morphological galaxy classification. The questions answered by web users from the Galaxy Zoo [16], a project built by astronomers to annotate large number of galaxy images with the help of common web users, also follow the general Hubble Sequence classification scheme.

The most related work to ours comes from De La Calleja *et al.* [7]. They utilized the Principal Component Analysis (PCA) to rotate galaxy images according to their first principle component (PC) and crop them based on the first 2 PCs. After that, they used the coefficients of the top PCs (Eigengalaxies) as the feature descriptor. Using eigengalaxies, they implemented a galaxy image retrieval system with nearest neighbor search on a dataset with only 309 images. Furthermore, on the same dataset, they also used eigengalaxies for image classification [6]. Their method reduces the effect of background noise and retrieve images that are invariant to position, orientation and size, but it fails to capture fine-grained local information like spiral arms that could be useful in discriminating certain kind of galaxies. Besides, the nearest neighbor search is not suitable for a large-scale dataset.

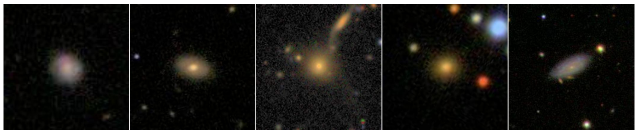


Figure 4. Example images from the dataset.

Goderya *et al*. [10] proposed a shape feature designed for galaxy images by binary thresholding. However, these shape features do not reveal much information of galaxies' inner structures and will fail in discriminating different types of galaxies with similar outer shapes. In addition, their feature is very sensitive to the choice of threshold value. For a galaxy image with low resolution and high noises, the thresholding could make a spiral galaxy looks like elliptical.

Recently, Davis and Hayes [5] did good work in quantitively describing spiral galaxy structures. They proposed a method that described spiral galaxy structures automatically as a set of arcs that fit spiral arm segments. The images they used are also came from the SDSS dataset. They chose $29,250$ galaxy images with sufficient human annotations for categories indicating visible spiral features from Galaxy Zoo project. Therefore, their method only works well for spiral galaxies that are viewed in good angles with relatively high resolution and low noise. For a spiral galaxy viewed edge-on or other kinds of galaxies such as elliptical or irregular, their method will not get reliable descriptions. The success in describing spiral galaxies shows that their method is complementary to ours and could be used to enhance the performance of spiral galaxy retrieval.

As for image retrieval, a trivial way is to use brute force nearest neighbor search. However, for a large-scale dataset, a good approximation method to trade-off between efficiency and accuracy is needed. Traditional approximation methods like KD-Trees [2] do not work well for data in high dimensional space and are sensitive to data distributions. Recently, hashing-based methods such as Locality Sensitive Hashing [3, 9] and Kernelized Locality Sensitive Hashing [14] become increasingly popular in large-scale image retrieval. The hashing-based methods not only achieve good performance experimentally but also have theoretical guarantee of sub-linear time in approximating nearest neighbor search.

3. Spatial-Color Layout Feature

3.1. Galaxy Detection and Alignment

The galaxy images we used come from the Galaxy Zoo project [16], which contains a subset of the Sloan Digital Sky Survey (SDSS) dataset. We used $283,971$ galaxy images and web user annotations corresponding to each image from the dataset. Figure 4 shows 5 randomly picked example images from our dataset.

In general, taking digital astronomical photographs needs high ISO (*i.e*. high film speed) and long exposure time, which introduces significant amount of dark-current noises [13]. Since astronomical images are noisy and contain lots of irrelevant background galaxies and stars, we need a good galaxy detector to remove the background and focus on the central galaxy. In addition, we also want to retrieve similar galaxies invariant to their scales and rotations, thus we need to align and crop all images so that all central galaxies share the same center and orientation.

Our galaxy detection and alignment approaches can be summarized in Figure 5. Firstly, for a given galaxy image (Figure 5(a)), we convert the image from RGB color space to gray-scale and use a edge-preserving bilateral filter [27] to remove noises from the gray-scale image. Secondly, we use Otsu's method [22] to adaptively find the threshold and segment the original image into a binary image (Figure 5(b)). After that, we fill holes in the binary image based on morphological operations [25] (Figure 5(c)). Then, we find the biggest connected component located in the center as our target galaxy. For the central galaxy in the original image, we calculate its center and orientation, where the center is defined as the center of mass and the orientation is defined as the direction that minimizes the second-moment of inertia (Figure 5(d)). We treat a pixel as a particle and the corresponding pixel value as its mass. The image was rotated around the center so that the orientation of the galaxy is pointed to the right horizontally. Finally, we calculate the radius of the minimum circle that encloses the whole central galaxy located at the center, on which we impose our spatial layout sections and extract our feature (Figure 5(e)).

3.2. Spatial-Color Layout Descriptor

Before introducing the proposed Spatial-Color Layout descriptor for galaxy images, we first discuss the intuitions of our descriptor from the perspective of Astronomy.

Studies from Astronomy show that a more massive star has shorter lifetime and bluer light emission compared to a less massive star [8]. Because of this, in general, younger stars are bluer and older stars are redder (younger means a massive star that burns very quickly and hence has shorter lifetime). Based on the Hubble Sequence, recent research in Astronomy shows that galaxy evolution is determined by a delicate balance between gas accretion and merging [26]. Nowadays, astronomers have reached a consensus on galaxy evolution. In general, galaxies evolve from spirals to ellipticals, which are believed to be the remnants of major merger events of spiral galaxies [1]. For a typical spiral galaxy, since star formation is inactive around the nucleus, stars at this location are basically older and redder (because most young blue stars died out quickly and no newborn stars generated, red stars with long lifetime will dominate). On the other hand, because star formation is very active in spi-

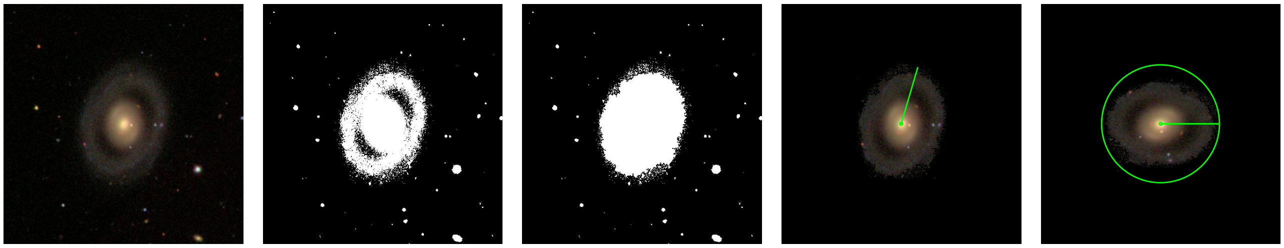


Figure 5. Galaxy detection and alignment steps. From left to right: original image; binary image after thresholding; binary image after filling holes; central connected component with its center and orientation; image after detection and alignment, the feature will be extracted from the circular area.

ral arms or marginal areas of the spiral galaxy, stars in spiral arms are younger and bluer. However, for an elliptical galaxy, there is almost no star formation so it consists mainly of red and yellow old stars [8].

Based on the knowledge of galaxy formation and evolution, we notice that a good discriminative feature for galaxy images should not only capture the global and local shape information, but also capture local color distributions. In addition, as a feature descriptor, it should also be robust to noise and invariant to scale and rotation. Thus, a good feature for galaxy images should have 4 properties:

- Captures both global and local shape information of the galaxy.
- Captures local color distributions of the galaxy.
- Robust to noise and changing of background.
- Scale and rotation invariant for the galaxy.

In light of this, we propose a novel feature that has all desired properties we mentioned above. The essence of our feature descriptor is concatenated color histograms extracted from a designed spatial layout (3-level circular regions), which is illustrated in Figure 6. So we call our feature a color-spatial layout feature.

Firstly, for a given galaxy image, we apply feature detector mentioned in Section 3.1 to detect and find the center and radius of the minimum circle that covers the central galaxy. Then, on the detected area, we impose a spatial layout that contains circular regions with 3 levels and 8 sections per level (Figure 6). The reason why we choose 3 levels is that a typical galaxy usually has a nucleus (bulge in center) and a disk or halo around it, so we want to build corresponding levels for the nucleus, the margin of the disk or halo, and the region in between respectively. In each level, we use 8 sections to capture the directional information from aligned galaxy images. After that, for each section, we extract a 16×3 dimension normalized color histogram (we tried RGB and Lab color space, the RGB performed better) and concatenate all 24 histograms as our feature. The dimension of the proposed feature is $16 \times 3 \times 24 = 1152$.

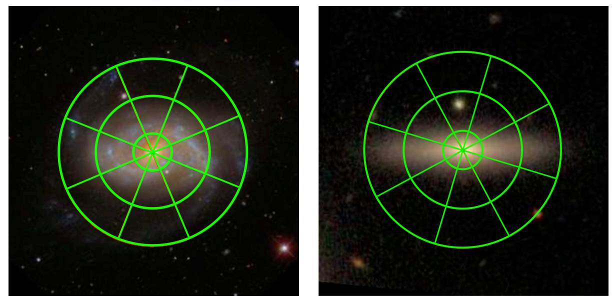


Figure 6. A spiral (left) and an elliptical (right) galaxy image after detection and alignment and their corresponding spatial layout where we extract our feature. The ratio of radii we used for the inner layer, middle layer and outer layer is $0.2 : 0.6 : 1.0$.

Next, we will use Figure 6 as an example to demonstrate that by capturing local color distributions of different regions in our spatial layout on a galaxy image, our feature descriptor preserves both local and global shape of the galaxy. Figure 6(a) and 6(b) illustrate a spiral and an elliptical galaxy image after detection and alignment respectively and their corresponding circular regions. Because of the simplicity of the background, pixels with low values will dominate those sections on background areas, for instance the uppermost three sections in the outer level of the elliptical galaxy (Figure 6(b)). Thus, the distribution of pixel values captures the global shape of a galaxy. For non-background sectors, the distribution of low-value pixels can also preserve the local shape of a galaxy. For example, the color distribution pattern in the outer level sections of Figure 6(a) encodes the structure of spiral arms. From the rightmost section in Figure 6(b), we can know the local curvature of the elliptical galaxy through the distribution of pixel values (by the proportion of the number of pixels with high values and those with low values). Furthermore, in Figure 6(a), the significant difference of color distribution between the inner layer and two outer layers (*i.e.* inner layer is redder and outer layer is bluer) implies the spiral shape with high confidence from the perspective of Astronomy.

To sum up, the feature detector removes the background noises and makes it robust to changing of background; the rotation and cropping in alignment makes our feature scale and rotation invariant; the local color distributions preserve both global and local shape, which is important to discriminate galaxy morphologically. Since the proposed feature descriptor uses statistical color information from regions, it is insensitive to noise pixels.

Therefore, the proposed feature has all the four ideal properties we mentioned earlier and is especially suitable for galaxy images.

4. Retrieval Method

After feature extraction, the problem is how to do fast image retrieval in such a large-scale dataset. Among those methods mentioned in Section 2, we are especially interested in Kernelized Locality Sensitive Hashing (KLSH) [14] for two reasons. Firstly, since our feature is a histogram based feature, KLSH enables us to use different nonlinear kernels such as Chi-Square, Histogram Intersection and Jenson-Shannon Divergence to compute the similarity between two features more accurately. Secondly, the efficient sub-linear time nearest neighbor approximation of KLSH is ideal for us to handle a large-scale dataset with more than $280,000$ images.

KLSH computes hash functions in kernel space based on a subset of p samples from a dataset containing n samples and $p = O(\sqrt{n})$ will guarantee a sub-linear search time [14]. We use 300 hash bits and choose $p = 100$ for our dataset with $n = 283,971$. We test on three different kernel functions [28] using KLSH:

- Chi-Square Kernel:

$$\mathcal{K}(x, y) = 1 - \frac{1}{2}\sum_{i=1}^{d}\frac{(x_i - y_i)^2}{x_i + y_i}, \tag{1}$$

- Histogram Intersection Kernel:

$$\mathcal{K}(x, y) = \sum_{i=1}^{d}\min(x_i, y_i), \tag{2}$$

- Jenson-Shannon Divergence Kernel:

$$\mathcal{K}(x, y) = \sum_{i=1}^{d}\Big(\frac{x_i}{2\log(\frac{x_i+y_i}{x_i})} + \frac{y_i}{2\log(\frac{x_i+y_i}{y_i})}\Big), \tag{3}$$

where $x, y \in \mathbb{R}^d$ are the two histogram features to be compared and $\sum_{i=1}^{d} x_i = \sum_{i=1}^{d} y_i = 1$. For the proposed feature, $d = 1,152$. $\mathcal{K}(x, y)$ is the non-linear similarity between x and y.

The comparison results of three kernels above will be shown in Section 5.

5. Experimental Results

A practical issue would be how to evaluate our proposed feature. We noticed that in order to label galaxy images, astronomers have launched a project called Galaxy Zoo [16]. This project invites the general public web users to visually inspect and annotate nearly one million galaxy images via the internet. Comparison results on a subset of SDSS show that the user labeled annotations from Galaxy Zoo are quite consistent with those classified by professional astronomers. Thus the annotation data provides a robust morphological catalogue [16], which could be used to build a meaningful evaluation standard for different galaxy image retrieval methods. We will evaluate the proposed feature based on the retrieval performance.

On the Galaxy Zoo website, web users are asked a sequence of hierarchical questions to describe a galaxy. There are 11 hierarchical questions and 37 associated answers in total. For example, the question "How prominent is the central bulge, compared to the rest of the galaxy?" has 4 possible answers: "No bulge", "Just noticeable", "Obvious" and "Dominant", from which a web user need to select one answer for this question. Each image in the dataset has been described by approximately 30 to 50 web users.

We refined the annotations by removing unreliable users' answers using methods described in [15]. For each image, we collected answers from all reliable users and used a 37-length normalized histogram to represent the distribution of their answers. Since different answers may have different weights in deciding whether two images are similar or not, we used a Support Vector Machine (SVM) to automatically learn the weights for 37 answers. We asked 3 human labelers to annotate a training set containing $7,200$ similar and dissimilar galaxy image pairs, $3,600$ for each. Then we built a SVM classifier to learn the weights of 37 answers in deciding whether two galaxies are similar or not from the training set. We tried different kernels for SVM including Linear, Histogram Intersection, Chi-Square and Jensen-Shannon Divergence kernels. Among them, Histogram Intersection kernel achieves the most agreements on the labeled pairs, which is 95.8% of accuracy by 10-folds cross validation. Since the SVM classifier achieved performance very close to human annotators, we trained a SVM with Histogram Intersection kernel on all $7,200$ labeled pairs and use the SVM model to evaluate retrieval performances.

The criteria we used to evaluate our retrieval system is "mean Average Precision at K" ($mAP@K$) [19]. For all the top K images retrieved by our system, if a retrieved image and the query are judged as a similar image pair by our evaluation system then it is counted as a hit, otherwise it is a miss. The "Precision at cut-off K" ($P@K$) is calculated by:

$$P@K = \frac{\#hits}{K}. \tag{4}$$

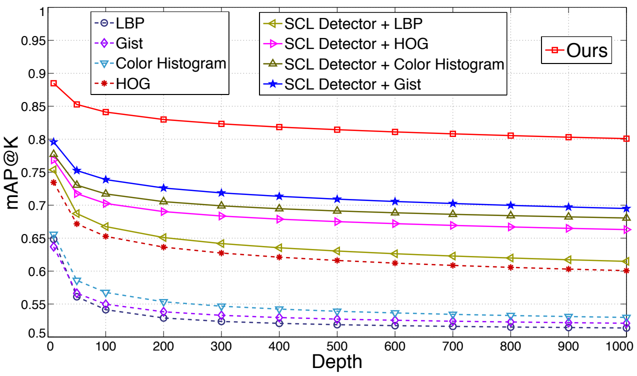


Figure 7. mAP comparison at different depths (K) between classic features and the proposed feature.

Then, for the given query, we can calculated the "Average Precision at K" ($AP@K$) by:

$$AP@K = \frac{\sum_{k=1}^{K} P@k \times Ind(k)}{P@K \times K}, \quad (5)$$

where $Ind(k)$ is an indicator function equaling 1 if the retrieved galaxy at rank k is counted as a hit, 0 otherwise. Finally, the $mAP@K$ is calculated by:

$$mAP@K = \frac{1}{N} \sum_{i=1}^{N} AP@K(i), \quad (6)$$

where N is the number of queries and $AP@K(i)$ is the $AP@K$ for the i-th galaxy image query. In our case, $N = 283,971$.

We compared the mAP at different depth (K) of the retrieval system using our proposed feature with those using other four classic features: Color Histogram, Local Binary Pattern (LBP), Gist and Histogram of Orientation (HOG). We used publicly available implementations for Gist, LBP and HOG from [21], [23] and [17] respectively. For color histogram, we used the normalized RGB color histogram of $64 \times 3 = 192$ dimension. The comparison results are shown in Figure 7. We used Histogram Intersection Kernel for our feature. For other features, the best results among using Linear, RBF, HIK, Chi-Square, JSD kernel are given. Our feature achieved 80.62% of $mAP@1000$. Compared with other features, clearly, our feature outperformed all four classic features by a large margin. In analyzing the performance of features, color histogram cannot preserve shape and cannot capture local information; LBP and HOG fail to capture color information and are not robust to noise; Gist cannot capture local and color information. They are all sensitive to background changes.

In addition, to illustrate the effectiveness of our feature detector, we also compared the performance of the above classic features before and after galaxy detection and alignment (steps showed in Figure 5). After galaxy detection means we only extract features from the circular region that encloses the whole galaxy after alignment (Figure 5(e)). As we can see from figure 7, the performances increased significantly after using our proposed Spatial-Color Layout Detector for all four features, which shows the effectiveness of the proposed feature detector.

Table 1. $mAP@1000$ using different kernels.

Kernel	Chi-Square	HIK	JSD
$mAP@1000$	80.08%	**80.62%**	77.93%

To show that local color is a more proper description for galaxy than local gradient, we obtained another feature from exactly the same spatial layout as of obtaining our feature descriptors. Instead of concatenating RGB color histograms, the feature is obtained by concatenating histograms of gradient. The $mAP@1000$ of this spatial-gradient layout feature is 67.63%, which is slightly higher than "SCL Detector + HOG" but still much lower than the performance of the proposed spatial-color layout feature. This confirmed the advantages of our feature descriptor.

The proposed feature is robust in terms of kernel selection. We can see from Table 1, the performance differs slightly for 3 different kernels. Furthermore, as we discussed in Section 3.2, our feature is invariant to scale and rotation. From Figure 8, we can see that using our feature, spiral galaxies with different scale and rotation to the query are successfully retrieved.

Figure 8. Scale and rotation invariant: the leftmost image is a query of spiral galaxy, the right five images are the top 5 galaxies retrieved by our system.

6. Other Applications

We believe there are three other major applications that could be very helpful in Astronomy using our proposed feature and retrieval system. Firstly, instead of galaxy retrieval, the proposed feature can also be used in automatic morphological galaxy classification. The second one is to discover unusual galaxies. If a query galaxy image is dissimilar from all its retrieved images, it is likely that the query galaxy is an unusual galaxy. Thirdly, for each galaxy in the dataset, the right ascension, declination and redshift can be retrieved from the database, as this information is stored in the SDSS archive. Thus, we could use them to locate a galaxy in celestial sphere [8] and visualize the distributions of a certain kind of galaxies in the universe. For example, Figure 9 and Table 2 show the retrieved merging galaxies and their locations using our system.

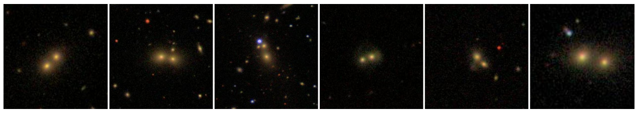


Figure 9. Finding and locating merging galaxies: the leftmost one is the query of merging galaxy, the right five images are the top 5 galaxies retrieved by our system. Their locations are given in Table 2.

Table 2. Finding and locating merging galaxies: right ascension (in degree), declination (in degree) and redshift bin of six merging galaxy images are shown in figure 9. This information can be used to visualize the location of specific galaxies in 3D space.

Image	RA	Dec	Red Shift Bin
Query Image	142.198	12.935	12
Top 1	120.563	29.395	11
Top 2	235.767	5.772	1
Top 3	145.653	12.142	12
Top 4	151.509	1.201	14
Top 5	179.979	0.617	9

7. Conclusions and Future Work

We proposed a spatial-color layout feature that is specifically designed for galaxy images. By combining the proposed feature with KLSH, an efficient galaxy image retrieval system has been built. On an Intel 3.2 GHz Quad-core PC, our system needs only 35 milliseconds to rank-order all images according to relevance to a given query in a dataset containing hundreds of thousands of images. We also proposed an evaluation system using web user annotations from the Galaxy Zoo project, which shows the benefits of the proposed feature, including being scale and rotation invariant, efficient, and more accurate in describing galaxies compared to other traditional features.

As future work, we plan to leverage our work in other Astronomical applications, including large-scale galaxy morphology classification and semantic description, using our proposed feature.

8. Acknowledgements

This work is an extended version of our course project in EECS 6890 Visual Recognition and Search at Columbia University [1] taught by Dr. Rogerio Feris and Dr. Liangliang Cao. We would like to thank them for their great helps and suggestions in this work. In addition, We also thank Professor Mary Putman at Department of Astronomy, Columbia University for her helps in understanding galaxy formation and evolution; Mr. Jie Feng, Mr. Joseph Ellis and Mr. Brendan Jou for kindly providing us a refined version of the Galaxy Zoo user annotations.

[1] http://rogerioferis.com/VisualRecognitionAndSearch2013

References

[1] J. E. Barnes and L. Hernquist. Dynamics of interacting galaxies. *Annual review of astronomy and astrophysics*, 1992. 3

[2] J. L. Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 1975. 3

[3] M. S. Charikar. Similarity estimation techniques from rounding algorithms. In *STOC*, 2002. 3

[4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005. 1

[5] D. R. Davis and W. B. Hayes. Automated quantitative description of spiral galaxy arm-segment structure. In *CVPR*, 2012. 3

[6] J. De La Calleja and O. Fuentes. Machine learning and image analysis for morphological galaxy classification. *Monthly Notices of the Royal Astronomical Society*, 2004. 2

[7] J. de la Calleja, O. Fuentes, A. Lopez-Lopez, and L. E. Erro. Content-based retrieval of astronomical images. 1, 2

[8] R. Freedman and W. J. Kaufmann. *Universe*. WH Freeman, 2007. 2, 3, 4, 6

[9] A. Gionis, P. Indyk, R. Motwani, et al. Similarity search in high dimensions via hashing. In *Proceedings of the international conference on very large data bases*, 1999. 3

[10] S. N. Goderya and S. M. Lolling. Morphological classification of galaxies using computer vision and artificial neural networks: A computational scheme. *Astrophysics and space science*, 2002. 3

[11] E. Hubble. The classification of spiral nebulae. *The Observatory*, 1927. 2

[12] E. P. Hubble. Extragalactic nebulae. *The Astrophysical Journal*, 1926. 2

[13] J. R. Janesick. *Scientific charge-coupled devices*. SPIE press Bellingham, WA, 2001. 3

[14] B. Kulis and K. Grauman. Kernelized locality-sensitive hashing for scalable image search. In *ICCV*, 2009. 2, 3, 5

[15] C. Lintott, K. Schawinski, S. Bamford, A. Slosar, K. Land, D. Thomas, E. Edmondson, K. Masters, R. C. Nichol, M. J. Raddick, et al. Galaxy zoo 1: data release of morphological classifications for nearly 900 000 galaxies? *Monthly Notices of the Royal Astronomical Society*, 2011. 5

[16] C. J. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M. J. Raddick, R. C. Nichol, A. Szalay, D. Andreescu, et al. Galaxy zoo: morphologies derived from visual inspection of galaxies from the sloan digital sky survey? *Monthly Notices of the Royal Astronomical Society*, 2008. 2, 3, 5

[17] O. Ludwig, D. Delgado, V. Gonçalves, and U. Nunes. Trainable classifier-fusion schemes: an application to pedestrian detection. In *ITSC*, 2009. 6

[18] R. Lupton, J. E. Gunn, Z. Ivezic, G. R. Knapp, S. Kent, and N. Yasuda. The sdss imaging pipelines. *arXiv preprint astro-ph/0101420*, 2001. 1

[19] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to information retrieval*, volume 1. Cambridge University Press Cambridge, 2008. 5

[20] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *TPAMI*, 2002. 1

[21] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *IJCV*, 2001. 1, 6

[22] N. Otsu. A threshold selection method from gray-level histograms. *Automatica*, 1975. 3

[23] M. Pietikäinen. Image analysis with local binary patterns. In *Image Analysis*. Springer, 2005. 6

[24] S. Sergyan. Color histogram features based image classification in content-based image retrieval systems. In *Applied Machine Intelligence and Informatics*, 2008. 1

[25] P. Soille. *Morphological image analysis: principles and applications*. Springer-Verlag New York, Inc., 2003. 3

[26] M. Steinmetz and J. F. Navarro. The hierarchical origin of galaxy morphologies. *New Astronomy*, 2002. 3

[27] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *ICCV*, 1998. 3

[28] A. Vedaldi and A. Zisserman. Efficient additive kernels via explicit feature maps. *TPAMI*, 2012. 5

[29] R. E. Williams et al. The hubble deep field: Observations, data reduction, and galaxy photometry. *arXiv preprint astro-ph/9607174*, 1996. 1