Mathematical and Computational Applications

MDPI

Article

A Sparse Representation Algorithm for Effective Photograph Retrieval

Hong-Bo Zhang *, Qing Lei, Bi-Neng Zhong, Ji-Xiang Du * and Duan-Sheng Chen

Department of Computer Science and Technology, Huaqiao University, Fujian 361021, China; leiqing@hqu.edu.cn (Q.L.); bnzhong@hqu.edu.cn (B.-N.Z.); dschen@hqu.edu.cn (D.-S.C.)

* Correspondence: zhanghongbo@hqu.edu.cn (H.-B.Z.); jxdu@hqu.edu.cn (J.-X.D.); Tel.: +86-136-969-26905 (H.-B.Z.)

Academic Editor: Fazal M. Mahomed
Received: 17 July 2016; Accepted: 11 January 2017; Published: 17 January 2017

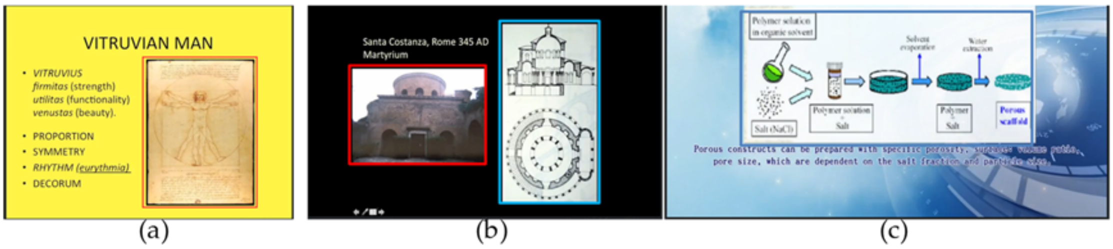
Abstract: Searching through information based on a photograph, which may contain graphics and images, has become a popular trend, such as in electronic books, journals, and products. Although many context-based methods have been proposed to retrieve images, most work focuses on selecting appropriate features for different objects. In the present study, we apply sparse representation to simultaneously retrieve image and graphics from a photograph. The sparse vector can be regarded as the similarity between the query photograph and dataset. The image with the largest entry (or several largest entries) can be assigned as the retrieved result. In the sparse representation framework, the common image features are used. Experimental results demonstrate that if the similarity vector in photograph retrieval is sparse, feature extraction is no longer critical. Compared with similar works in photograph retrieval, the proposed method has better retrieval accuracy.

Keywords: photograph retrieval; sparse representation; similarity; sparse vector; feature extraction

1. Introduction

Photograph retrieval is a key component in many Web and e-business applications such as searching electronic books, journals, and products, as well as related information retrieval, e.g., mobile visual research [1], e-learning systems [2], and modality recognition for medical images [3]. With the popularization of mobile devices, photographs have become a new and important method to record information. Photograph retrieval is critical in mobile visual research. In an e-learning system, students can learn remotely by watching videos, and important concepts and data are often presented using graphs and images. In class, students record information using mobile cameras. Students can use graphics and images in lecture videos as query to retrieve knowledge from an electronic database. Especially, these are many graphs rather than nature images in the textbooks and materials. Thus, an effective photograph-retrieval system is necessary. Figure 1 shows some examples of photographs from a presentation file in video form.

Although many content-based retrieval methods have been developed for images and graphs, few are simultaneously designed for both images and graphs. In our previous work [4,5], an image and graph classifier was proposed using the difference in the entropy of grayscale distribution for adaptive photograph retrieval. A photograph is regarded as a pixel-based feature that contains a histogram with an oriented gradient. For images and graphics, the difference in the low-level feature is used. The similarity between the query photograph and each database prototype is evaluated by feature matching using different distance measures. However, because the representation of graphics is simple and similar among graphics, they are more challenging. Figure 2 shows a photograph-retrieval problem.



VITRUVIAN MAN

- *VITRUVIUS* *firmitas* (strength) *utilitas* (functionality) *venustas* (beauty).
- PROPORTION
- SYMMETRY
- *RHYTHM (eurythmia)*
- DECORUM

Santa Costanza, Rome 345 AD Martyrium

Porous constructs can be prepared with specific porosity, surface: volume ratio, pore size, which are dependent on the salt fraction and particle size.

Figure 1. Examples of photographs from a presentation file in a video form. The images are marked by red rectangles, and the graphs are marked by blue rectangles. (**a**): example of presentation file only containing image; (**b**): example of presentation file containing image and graph; (**c**): example of presentation file only containing graph.

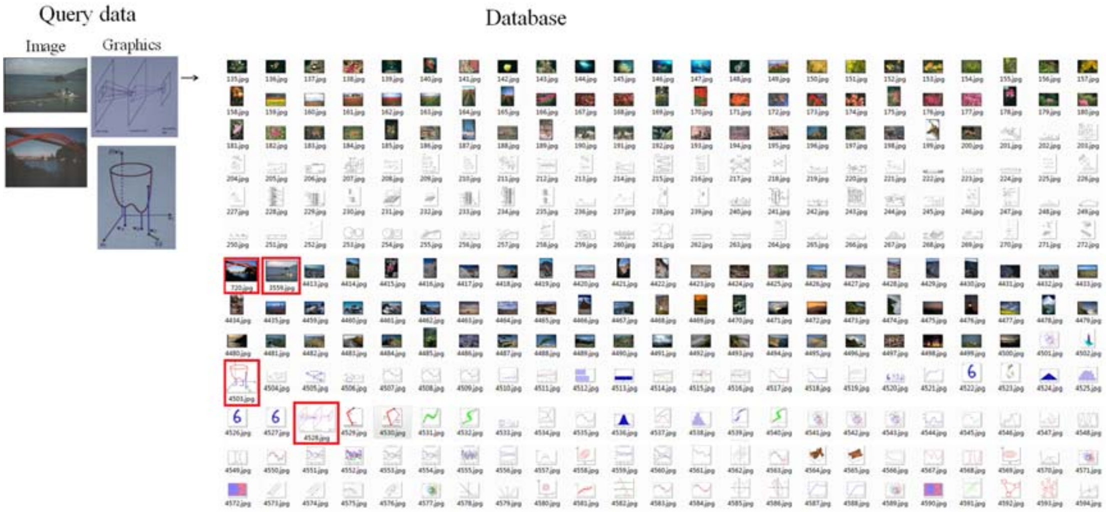


Figure 2. Examples of photograph retrieval. The red rectangles are the query results.

Many low-level features have been used in context-based image-retrieval methods that contain pixel- and contour-based features. Pixel-based features perform computation using all pixels or edge pixels in an image. Contour-based features perform computation using lines or curves. Popular pixel-based features include the scale-invariant feature transform [6], pyramid histogram of oriented gradients (PHOG) [7], GIST feature [8] which is the abstract representation of the image, shape context [9], and structured local binary Haar pattern [10]. Typical contour-based features include triangle area representation [11], local structure [12], inner-distance shape [13], and local neighborhood structure shape [14]. After feature extraction, a near-neighbor search is applied for retrieval.

In contrast to these works, we propose a more robust method based on sparse representation that does not need to classify images and graphics and chooses different features. The main idea behind the sparse representation in photograph retrieval is that given a sufficiently diverse database, the query photograph can be well represented as a sparse linear combination of the database. In the database, the query photograph is the only data. This condition would naturally encode the similarity information between the query and database into a sparse representation. This is a natural concept of sparse representation. To reduce the time complexity in solving the sparse representation, random projection as proposed by Foroughi et al. [15] is used as a dimensionality reduction method. Because more information is involved, the proposed method is superior to our previous method, as demonstrated by experiments. If sparsity is the intrinsic quality of photograph retrieval, the feature is no longer critical. In particular, for graphics retrieval, sparse representation also performs better.

The contributions of our work are threefold:

(1) In the proposed sparse representation framework, the sparse vector is firstly used to measure the similarity between the query photograph and dataset, rather than feature distance.

(2) Unlike existing methods, which use feature-based sparse representation to build a dictionary for content-based image retrieval, we directly use the image features as a dictionary. If sparsity is the intrinsic quality of photograph retrieval, the feature is no longer critical.

(3) The experimental result show that the proposed method is an effective, robust image and graph classification method and provides more accurate results.

The remainder of this paper is organized as follows. Section 2 introduces the related works. Section 3 describes the algorithms in which the proposed method is based. Section 4 presents and discusses the experimental results. Section 5 concludes this paper.

2. Sparse Representation

The sparse representation of high-dimensional data has generated increasing interest in machine learning and computer vision, especially for image classification and object recognition [16–19]. The basic idea of sparse representation is that the input signal (image) can be expressed as a linear combination of an overcomplete dictionary and sparse vector. The key problems in sparse representation are how to construct the overcomplete dictionary and how to solve the sparse vector. Popular technologies to construct the dictionary include the method of optional directions, k-singular value decomposition (K-SVD), discrete cosine transform, and online learning algorithm. The sparse vector can be effectively computed by greedy methods or optimization. Typically, sparse representation is regarded as an l_0-minimization problem.

In recent works, variations and extensions in solving sparse representation have been proposed for many computer-vision tasks based on compressive sensing theory. Ortiz et al. [20] proposed a novel linearly approximated sparse representation-based classification algorithm that uses linear regression to perform sample selection for l_1-minimization. Yang et al. [21] proposed a new robust sparse coding by modeling the sparse coding as a sparsity-constrained robust regression problem. Robust sparse coding seeks the maximum likelihood estimation solution of the sparse coding problem. Yigang et al. [22] proposed a novel method based on sparse and low rank decomposition for linearly correlated images. The challenging optimization problem is reduced to a sequence of convex programs that minimize the sum of l_1-norm and nuclear norm of two component matrices.

In recently years, many approaches have been proposed for content-based image retrieval. Most works focus on feature representation, and feature-based sparse representation is used to build dictionary learning. In [23], a high-order feature is built to improve the retrieval accuracy. In [24], a clustering method using dictionary learning is proposed to group large medical dataset. The sparse representation based method is proposed to learning dictionary via K-SVD decomposition. In [25], the iterative discrete wavelet transform is proposed to extract features and sparse representation is used to build the dictionary. Unlike these methods, we directly use the image features as a dictionary and apply the sparse vector as the similarity between the query photograph and dataset.

3. Photograph-Retrieval Method Based on Sparse Representation

The proposed algorithm is shown in Figure 3. First, a holistic representation of the photograph is needed for the query image and database. The query image is collected from electronic books, journals, and representation files. In the practical application, the query usually contains noise, illumination variation and shadow. The dictionary is set up by a combination of the image features in the database. The sparse representation of the query image is solved by l_1-minimization. Finally, we obtain the query result image by searching the maximum element in the sparse vector.

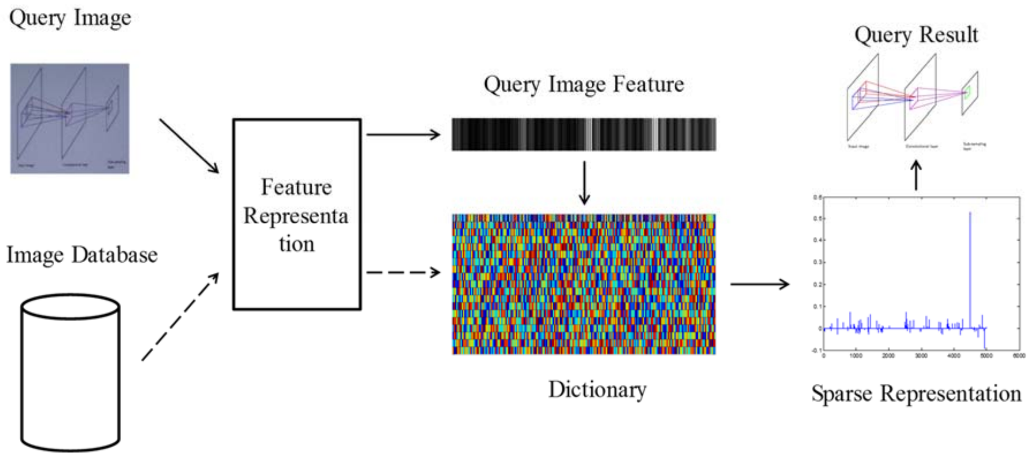


Figure 3. Framework of sparse representation-based photograph-retrieval method.

3.1. Image and Graph Feature

Many holistic features can be used for photograph retrieval. To verify that the feature is no longer critical when the photograph retrieval is regarded as sparse representation, the GIST and PHOG features were applied in the proposed method. The GIST feature divided the image into 4×4 regions. A set of Gabor filters with different frequencies and orientation was used to compute the region feature. The image feature was the combination of these region features.

The PHOG feature represented the image using the histogram of the gradient orientation. The image was divided with the spatial pyramid. The parameters of the feature extraction are presented in the experiment section.

3.2. Photograph Retrieval via Sparse Representation Algorithm

3.2.1. Sparse Representation for Photograph Retrieval

Let us suppose that the image database contains n images. From the perspective of image classification, each image can be regarded as an independent class. The image database can be regarded as set $D = \{x_i\}$, where $x_i \in R^m$ is the feature vector of the image and m is the feature dimension. Given any query image $q \in R^m$, this query can be expressed as a linear approximation of the database.

$$q = A\alpha \in R^m \tag{1}$$

where $A = [x_1^T, \ldots, x_i^T, \ldots, x_n^T] \in R^{m*n}$ is the dictionary that concatenates all the images in the database. $\alpha \in R^n$ is a coefficient vector whose entries are zero except those associated with the correct or similar images. In other words, vector α is the sparse representation of the query image. Naturally, the solution to Equation (1) requires the minimum of $\|\alpha\|_0$, which is found by solving the following optimization problem:

$$\begin{aligned} &a^* = \arg\min\|\alpha\|_0 \\ &s.t. \quad q = A\alpha \end{aligned} \tag{2}$$

Directly solving the sparse coefficient using Equation (2) is difficult. Under some conditions, we prove that the l_1-minimization problem is equivalent to l_0-minimization. The optimization problem of Equation (2) can be transformed into the following optimization problem:

$$\begin{aligned} &a^* = \arg\min\|\alpha\|_1 \\ &s.t. \quad \|A\alpha - q\|_2^2 \leq \varepsilon \end{aligned} \tag{3}$$

where ε is the noise level. Equation (3) can be transformed as follows using the Lagrangian multiplier:

$$a^* = \arg\min \|A\alpha - q\|_2^2 + \lambda\|\alpha\|_1 \tag{4}$$

Equation (4) can be solved by many optimization algorithms. In our work, the strategy proposed in [15] is applied.

3.2.2. Query Result Determination

Once sparse vector α^* is determined by Equation (4), the query result is easily assigned to a query image. Ideally, only one nonzero element exists in α^* and corresponds to the correct search image. However, because of noise, modeling error, and similar images, especially in graph retrieval, some small nonzero elements exist in α^* associated with multiple images in the database, as shown in Figure 4. A design classifier using a different algorithm is one method of solving this problem. In our work, we assign the largest response image in α^* to the query. Here, we learn that $\alpha^* = [a_1, \ldots, a_i, \ldots, a_n]$, where α_i means the response (or similarity) between the ith image in the database and query. The query result is the ith image that maximizes α_i.

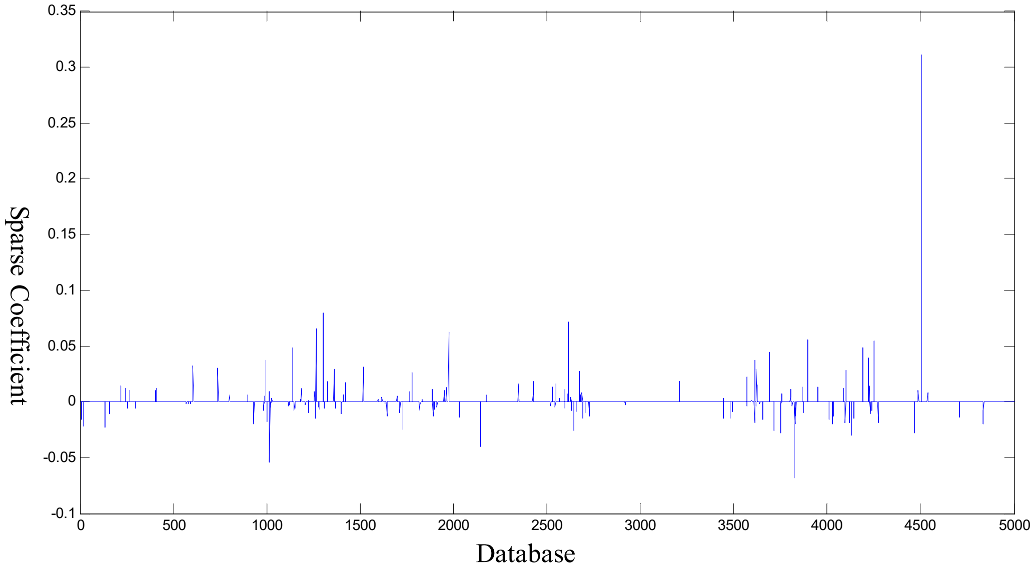


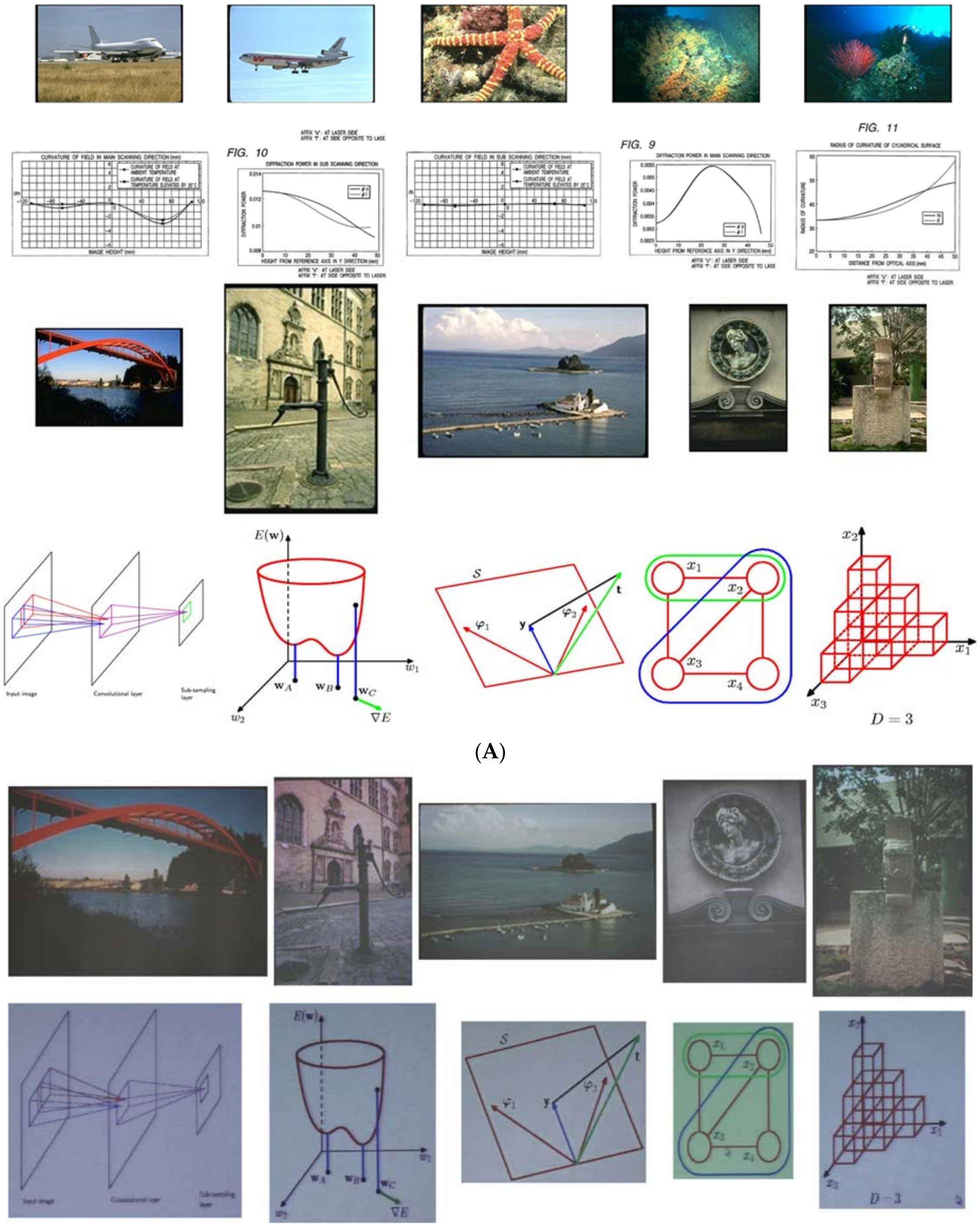
Figure 4. Example of sparse vector α^*.

3.2.3. Computational Complexity

The intensive computation of the proposed method is to solve the l_1-minimization problem defined in Equation (4). The l_1-minimization problem is formulated as a linear approximation and solved by classical methods in convex optimization. In [15], Foroughi et al. drew extensively on the survey which compared the performance of different l_1-minimization method for sparse optimization. According these comparison, we also select the Homotopy and Dual Augmented Lagrangian Methods (DALM) for fast l_1-minimization. The computational time is $O(m^2 + mn)$, where m is feature dimension and n is the number of the images in the dataset.

4. Experimental Results

To evaluate the performance of the proposed method, we employed the 5000-database proposed by [4]. The database has 5000 data that include 2500 images and 2500 graphics (drawing or diagrams). The query set has 100 data that include 50 images and 50 graphics. Figure 5 shows some examples of the 5000-database and query set.



(**A**)

(**B**)

Figure 5. Examples of the database prototypes. (**A**) 5000-database. (**B**) Query dataset.

4.1. Parameter Setting in the Experiment

To calculate the GIST feature, the image was divided into 4×4 regions. The Gabor filter was applied in each region. The Gabor filter contained four orientations, and each orientation has eight scales. Finally, the dimension of the GIST feature was 512.

To calculate the PHOG feature, the image was divided using the spatial pyramid structure. In the experiment, the pyramid levels were set to four, from $l = 0$ to $l = 3$, and contained 85 regions. In each region, the histogram of the gradient orientation was calculated using eight bins. Finally, the dimension of the PHOG feature was 680. The accuracy was used to evaluate the performance of the proposed method. When the original image of the query image was accurate, the query result was correct. The accuracy is calculated as follows:

$$Accuracy = \frac{numbers\_of\_correct\_query}{numbers\_of\_query\_set} \tag{5}$$

To make the evaluation of the proposed method reasonable, we also used precision and recall criteria as the performance measures of the content based image retrieval (CBIR) systems. A combination of the precision and recall criteria was used as performance measures in [25]. The precision and recall criteria are calculated as follows:

$$Precision = \frac{numbers\_of\_relevant}{total\_numbers\_of\_retrieval} \tag{6}$$

$$Recall = \frac{numbers\_of\_relevant}{total\_numbers\_of\_relevant\_in\_datasets} \tag{7}$$

where $P(0.5)$ is precision at 50% recall; $P(1)$ is precision at 100% recall.

4.2. Experimental Result Analysis

Here, some experiments are performed to prove the increase in the proposed method in terms of accuracy. To evaluate the effectiveness of the sparse representation for photograph retrieval, the nearest neighbor (NN) search is applied with image features as the basic line. Four different distance measures are available, which include the Euclidean, cosine, correlation, and Chebychev distances, and are used for the NN search. The experimental results are listed in Tables 1 and 2.

Table 1. Photograph-retrieval result with GIST feature in the 5000-database.

GIST	SR[1]	Euclidean-NN[2]	Cosine-NN[2]	Correlation-NN[2]	Chebychev-NN[2]
Image (50)	94.00%	80.00%	78.00%	84.00%	42.00%
Graphics (50)	94.00%	78.00%	80.00%	84.00%	76.00%
Total	94.00%	79.00%	79.00%	84.00%	59.00%

[1] SR: Sparse Representation; [2] NN: Nearest Neighbor.

Table 2. Photograph-retrieval result with PHOG feature in the 5000-database.

PHOG	SR	Euclidean-NN	Cosine-NN	Correlation-NN	Chebychev-NN
Image (50)	98.00%	68.00%	70.00%	70.00%	30.00%
Graphics (50)	88.00%	22.00%	22.00%	22.00%	84.00%
Total	93.00%	45.00%	46.00%	46.00%	16.00%

From the results, the accuracy of the sparse representation with GIST and PHOG features are 94.00% and 93.00%, respectively. It has a 10% increase compared with the best NN search with the GIST feature. Further, it has a 9% increase compared with the best NN search with the PHOG feature. The best NN search means the best query result using different distance measurements.

Table 3 lists the comparison of the proposed method with other related works. According to the comparison, the proposed method sees an improvement of over 10% compared with the other methods.

Table 3. Comparison results of the photograph retrieval under different methods.

Query Data	GIST-SR	PHOG-SR	[4]	[10]
Image (50)	94.00%	98.00%	90.00%	59.00%
Graphics (50)	94.00%	88.00%	74.00%	77.73%
Total	94.00%	93.00%	82.00%	68.81%

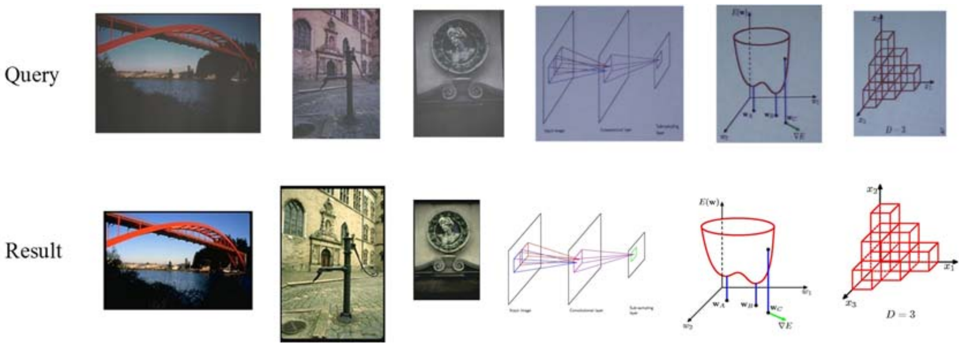
To verify the proposed method, we have compared the proposed method with the algorithm using sparse representation [25]. The comparison results of the precision and recall criteria are shown in Table 4. The best performance rate is $P(1) = 96.66\%$ and $P(0.5) = 97.34\%$, which is higher than that achieved using the algorithm proposed in [25].

Table 4. Comparison results of the retrieval system for the 5000-database.

Method	$P(1)$	$P(0.5)$
GIST-SR	96.66%	97.34%
PHOG-SR	96.16%	96.84%
[25]	95.74%	96.13%

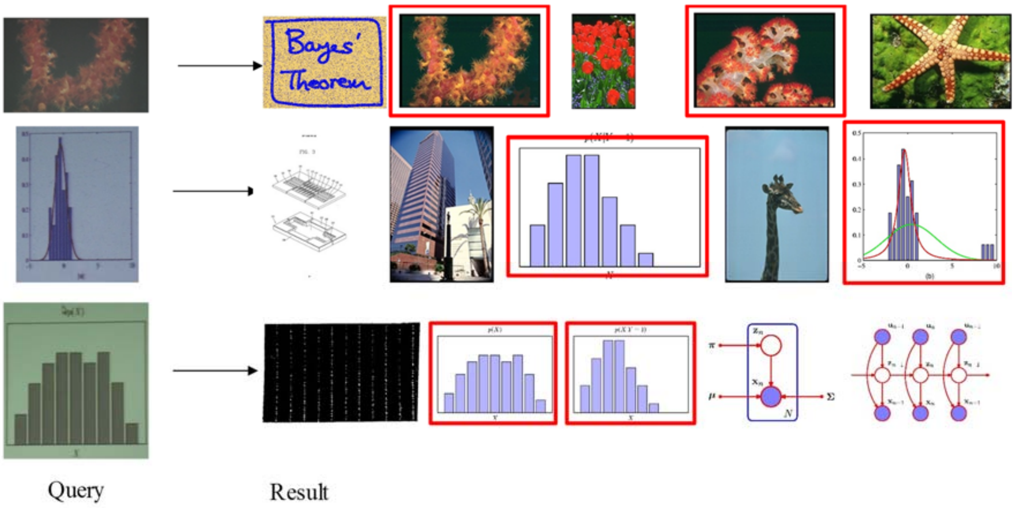
Figure 6 shows some query results. If we only count the nearest retrieval result which has the largest coefficient in the sparse vector, the precision with the GIST feature is 94% and the precision with PHOG is 93%. Figure 6A means the correct query results and Figure 6B is an example of an incorrect query and show more query responses.

To analyze the reason for the incorrect result, Figure 7 shows the sparse vector of some incorrect query photographs. In the experimental results, we choose the image which has the highest response value as the query result. However, for these incorrect results shown in Figure 6, we find that there are multiple high response values for query photograph as shown in Figure 7. Generally, the confidence of the query result with multiple high response values is lower. Ideally, the query result only has a high response value. In this respect, although there are some incorrect results, we also can infer that the result is not credible in the proposed method.



(**A**)

Figure 6. *Cont.*



(**B**)

Figure 6. Examples of the query result. (**A**) Correct query. (**B**) Incorrect query. The red rectangle shows the relevant result.

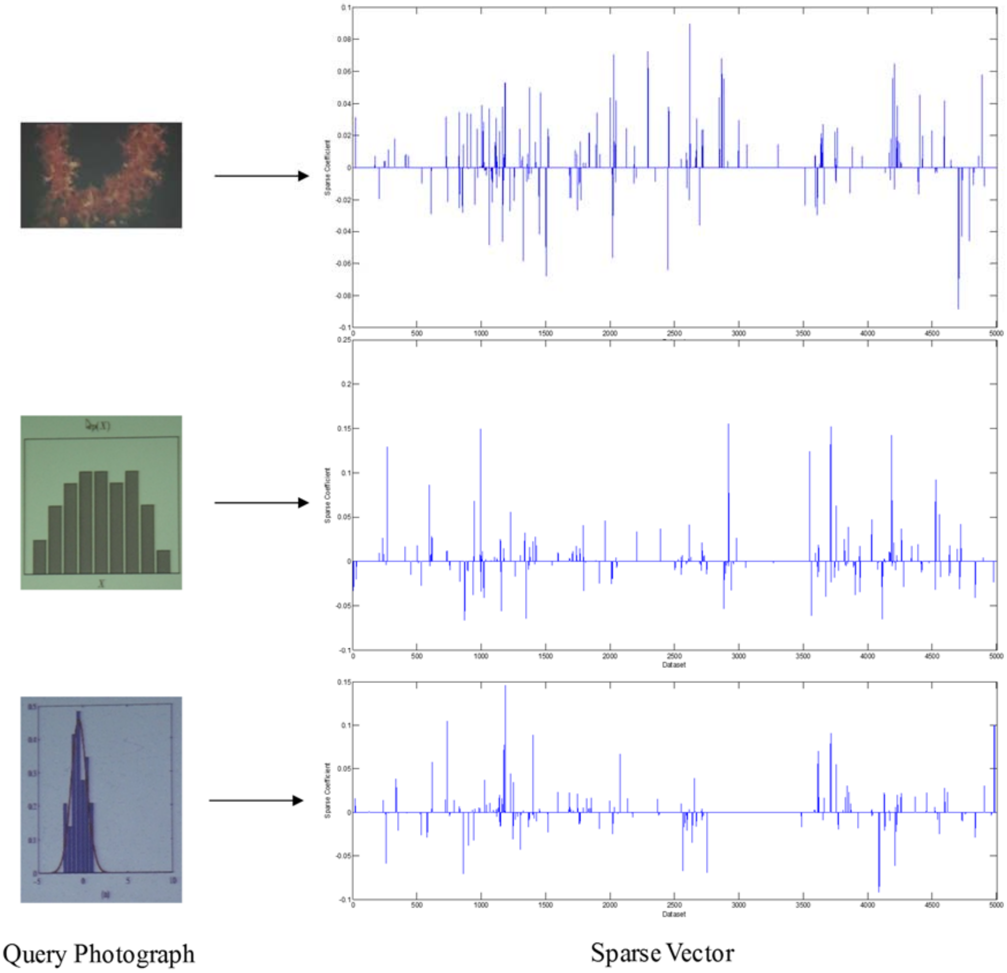


Figure 7. The sparse vector of the incorrect query.

Comparison of the GIST and PHOG features with the sparse representation reveals almost no difference in accuracy. However, for the NN search, the GIST feature performs better than the PHOG feature. From these results, the sparse representation is more robust with different features. In other words, when the photograph retrieval is sparse, the feature is not very critical.

On the other hand, in the experiment, the query image contains a variety of contents, such as landscape image, face image and flow diagram, etc. From the results, the proposed method is effective for these different contents.

Based on the analysis above, in the proposed method, the query image is regarded as the sparse linear approximation of the dataset, which is not dedicated to such kind of content. The proposed method is general enough for other kinds of contents, i.e., we are open to other off-the-shelf feature detectors.

5. Conclusions

In this work, a sparse representation-based method has been used to simultaneously retrieve images and graphics. Unlike exiting methods, which use feature-based sparse representation to build a dictionary for contend-based image retrieval, we directly use the image features as dictionary. The sparse vector is firstly applied to measure the similarity between the query photograph and the dataset. The experimental results demonstrate that the proposed method has better retrieval accuracy in photograph retrieval compared with similar works. The sparse representation was verified to be more robust with different features.

Acknowledgments: The authors would like to thank the anonymous reviewers for the valuable and insightful comments on the earlier version of this manuscript. The work was supported in part by the Nature Science Foundation of China (No. 61502182, 61673186, 11401231 and 61572205), by the Natural Science Foundation of Fujian Province of China (No. 2015J01253, 2015J01257, 2013J06014 and 11171040), by the Pilot Project of Fujian Province of China (No. 2015H0025), and by the Promotion Program for Young and Middle-aged Teachers in Science and Technology Research of Huaqiao University (No. ZQN-YX108).

Author Contributions: H.B.Z. and Q.L. conceived and designed the algorithm; B.N.Z. carried out the GIST feature; J.X.D. carried out the sparse representation algorithm. D.S.C. assisted with experimental results analysis. H.B.Z. wrote the paper. All authors read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ji, R.; Duan, L.; Chen, J.; Yao, H.; Yuan, J.; Rui, Y.; Gao, W. Location Discriminative Vocabulary Coding for Mobile Landmark Search. *Int. J. Comput. Vis.* **2012**, *96*, 290–314. [CrossRef]
2. Chio, J.-W.; Chen, S.-Y. Illustration Extraction from Video Streams. *J. Pattern Recognit. Res.* **2012**, *7*, 16. [CrossRef]
3. Dimitrovski, I.; Kocev, D.; Kitanovski, I.; Loskovska, S.; Džeroski, S. Improved medical image modality classification using a combination of visual and textual features. *Comput. Med. Imaging Graph.* **2015**, *39*, 14–26. [CrossRef] [PubMed]
4. Zhang, H.B.; Li, S.; Chen, S.; Su, S.Z.; Duh, D.; Li, S.Z. Adaptive photograph retrieval method. *Multimed. Tools Appl.* **2014**, *70*, 2189–2209. [CrossRef]
5. Li, S.A.; Chen, S.; Su, S.; Duh, D.; Li, S. Graphics/Image Retrieval Method. In Proceedings of the 2011 Conference on Technologies and Applications of Artificial Intelligence, Taoyuan, Taiwan, 11–13 November 2011.
6. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
7. Bosch, A.; Zisserman, A.; Munoz, X. Representing shape with a spatial pyramid kernel. In Proceedings of the 6th ACM International Conference on Image and Video Retrieval, Amsterdam, The Netherlands, 9–11 July 2007; Association for Computing Machinery: New York, NY, USA, 2007.
8. Oliva, A.; Torralba, A. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. J. Comput. Vis.* **2001**, *42*, 145–175. [CrossRef]

9. Belongie, S.; Malik, J.; Puzicha, J. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 509–522. [CrossRef]
10. Su, S.Z.; Chen, S.Y.; Li, S.Z.; Li, S.A.; Duh, D.J. Structured local binary Haar pattern for pixel-based graphics retrieval. *Electron. Lett.* **2010**, *46*, 996–998. [CrossRef]
11. Naif, A.; Kamel, M.S.; Freeman, G.H. Geometry-based image retrieval in binary image databases. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 1003–1013.
12. Chi, Y.; Leung, M.K.H. ALSBIR: A local-structure-based image retrieval. *Pattern Recognit.* **2007**, *40*, 244–261. [CrossRef]
13. Haibin, L.; Jacobs, D.W. Shape classification using the inner-distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 286–299.
14. Liu, R.; Wang, Y.; Baba, T.; Masumoto, D. Shape detection from line drawings with local neighborhood structure. *Pattern Recognit.* **2010**, *43*, 1907–1916. [CrossRef]
15. Foroughi, H.; Ray, N.; Zhang, H. Robust people counting using sparse representation and random projection. *Pattern Recognit.* **2015**, *48*, 3038–3052. [CrossRef]
16. Rigamonti, R.; Brown, M.A.; Lepetit, V. Are sparse representations really relevant for image classification? In Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, Colorado Springs, CO, USA, 20–25 June 2011.
17. Yang, J.; Wright, J.; Huang, T.S.; Ma, Y. Image super-resolution via sparse representation. *IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc.* **2010**, *19*, 2861–2873. [CrossRef] [PubMed]
18. Adler, A.; Helor, Y.; Elad, M. *A Shrinkage Learning Approach for Single Image Super-Resolution with Overcomplete Representations*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 622–635.
19. Matan, P.; Michael, E. Image sequence denoising via sparse and redundant representations. *IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc.* **2009**, *18*, 27–35.
20. Ortiz, E.G.; Becker, B.C. Face recognition for web-scale datasets. *Comput. Vis. Image Underst.* **2014**, *118*, 153–170. [CrossRef]
21. Yang, M.; Zhang, L.; Yang, J.; Zhang, D. Robust sparse coding for face recognition. In Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, Colorado Springs, CO, USA, 20–25 June 2011.
22. Peng, Y.; Ganesh, A.; Wright, J.; Xu, W.; Ma, Y. RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 2233–2246. [CrossRef] [PubMed]
23. Zhang, S.; Huang, Q.; Hua, G.; Jiang, S.; Gao, W.; Tian, Q. Building contextual visual vocabulary for large-scale image applications. In Proceedings of the International Conference on Multimedea 2010, Firenze, Italy, 25–29 October 2010.
24. Srinivas, M.; Naidu, R.R.; Sastry, C.S.; Mohan, C.K. Content based medical image retrieval using dictionary learning. *Neurocomputing* **2015**, *168*, 880–895. [CrossRef]
25. Mohamadzadeh, S.; Farsi, H. Content-based image retrieval system via sparse representation. *Iet Comput. Vis.* **2016**, *10*, 95–102. [CrossRef]