A Novel Graph Embedding Framework for Object Recognition

Mario Manzo[1], Simone Pellino[1], Alfredo Petrosino[1(✉)], and Alessandro Rozza[2]

[1] University of Naples Parthenope, Naples, Italy
petrosino@uniparthenope.it
[2] Research Team - Hyera Software, Coccaglio, Italy

Abstract. A great deal of research works have been devoted to understand image contents. In this field many well-known methods exploit Bag of Words (BoW) features describing image contents as appearance frequency histogram of visual words. These approaches have a main drawback, the location information and the relationships between features are lost. To overcame this limitation we propose a novel methodology for the Object recognition task. A digital image is described as a feature vector computed by means of a new graph embedding paradigm on the Attributed Relational SIFT Regions Graph. The final classification is performed by using Logistic Label Propagation classifier. Our framework is evaluated on standard databases (such as ETH-80, COIL-100 and ALOI) and the achieved results compared with those obtained by well-known methodologies confirm its quality.

Keywords: Image classification · Object recognition · Graph based image representation · Graph embedding

1 Introduction

In the last decade a great deal of research has been devoted to 3D object recognition. In order to capture distinctive details of the images, most of the image representation techniques leverage local features, such as SIFT [15] and HOG [18]. Unfortunately, local features would require to solve an assignment problem between every image pair, thus making it unfeasible to use them in real world scenarios. For this reason, a common strategy to integrate the local features into a fixed length global representation is to use the Bag Of Words approach. This technique is roughly composed of three steps: the local features extraction, the codebook generation and local features encoding, and the code pooling to generate the global image representation [5].

Despite the promising results achieved employing this kind of features, the main problem of this approach is due to the fact that the location and spatial information between local features are not considered. In order to solve this problem some research works exploit local and structural information of the image

L. Agapito et al. (Eds.): ECCV 2014 Workshops, Part IV, LNCS 8928, pp. 341–352, 2015.
DOI: 10.1007/978-3-319-16220-1_24

by employing graphs to model it in order to add some high level information (relations) to the low-level representation of the individual parts [19,30].

In this work, we propose an object recognition framework that is able to represent images in order to capture local and structural information. First of all an image is represented by the ARSG structure, as proposed in [16]. This structure encodes the SIFT features extracted from the image in a hierarchical fashion by considering both the individual image features and more global image regions. Pairwise relationships between features and regions are encoded in an incidence graph, which serves as a reduced representation for the entire image.

Originally, in [16] image comparison was performed through the direct comparison of the derived graphs. Since direct comparison to each image in the database becomes infeasible in this setting, we propose to extract a set of sample graphs and characterize each image through a feature vector (graph embedding), whose i-th coordinate is the similarity of ARSG graph of the given image to the i-th sample graph. This embedding paradigm, as proposed in [24], was efficiently experienced for non rigid scene recognition showing state-of-the-art results on datasets like SUN 397 [31] and KTH-IDOL 2 [10]. The main idea of this paper is to adapt this representation to the context of 3D object recognition in the presence of a potentially large image database. A semi-supervised learning approach, by means of a Logistic Label Propagation (`LLP`, [13]) algorithm, is adopted to accurately estimate the label values as the posterior probabilities. The results achieved on standard datasets compared with those obtained by well-known approaches confirm the quality of the proposed framework.

The paper is organized as follows: Section 2 summarizes the related works; Section 3 introduces the proposed object recognition framework; Section 4 describes the experimental results; Section 5 presents our conclusions and future works.

2 Related Works

The Object recognition is the task of finding and identifying objects in a video sequence or image. Given an image containing objects of interest and a set of labels, corresponding to a set of known models, the aim is to assign correct labels to regions that contains the objects of interest. This task is very difficult particularly if we consider the design of the recognition system. The main problems concern the object representations and their classifications. The goal is to emulate human system, which performs efficiently and dynamically the Object recognition task.

To capture distinctive details of the images, many image representation techniques leverage local features, such as `SIFT` [15] and `HOG` [18]. We recall that, unfortunately, the usage of local features as they are would require to solve an assignment problem between every image pair, thus making it unfeasible to use them in real world scenarios. For this reason, a common strategy to integrate the local features into a fixed length global representation is to use the Bag Of Words approach [5].

In [29] an M^{th} order tensor discriminant analysis approach for object categorization and recognition is described. This tensor approach avoids to transform 3D color images or 2D grayscale images into high dimensional feature vectors. The method represents a color image as a M^{th} order tensor and the original tensor objects are mapped into a low dimensional feature space where nearest neighbor classifiers and AdaBoost (hereinafter `DTROD-AdaBoost`) are employed to perform the final classification.

In [23] an approach to object recognition is described. It is based on matching of local image features. Precisely, the method recognizes objects under very different viewing conditions. The main idea concerns several affine-invariant constructions of Local Affine Frames (`LAF`s) for local image patches extraction. The robustness of the matching procedure is performed by giving multiple frames to each image region detected, and selecting the most discriminative ones. Matching score is estimated as the number of established local correspondences, without enforcing a global model consistency.

Despite the quality of the results obtained by employing the aforementioned approaches the location and spatial information between local features are not considered. In this context, graph structures can be a great help in order to reduce the gap between the location and spatial information of local features. Graphs are adopted in application domains where relations among data (edges) must be highlighted. Image processing [1,16], pattern recognition [7,11,25], and many other fields benefit from data graph representations and related manipulation algorithms.

The most used graph-based image representation is the Region Adjacency Graph (`RAG`) [27] in which a node represents a region of the image and an edge exists between two nodes if the underlying regions are adjacent. Despite this representation is widely used, there are other interesting alternatives in literature. Among them, in [9] a method for generic object recognition through graph structural expression using `SIFT` features is described. This approach creates a graph structure that connects the `SIFT` keypoints. This formulation reduces the computation complexity and, at same time, improves the detection performance.

In [7] a graph mining algorithm, called `gdFil`, is described. This work exploits two novel properties that allow to remove all duplicate candidates in Frequent Connected Subgraph (`FCS`) before support calculation. Support calculation task is addressed through a strategy based on embedding structures.

In [11] a graph mining framework called APproximate Graph Mining (`APGM`) is proposed. The framework is designed to identify approximate matched `FCS`s and to mine useful patterns from noisy graph database.

In [1] another graph mining algorithm for `FCS`s over undirected and labeled graph collections (called Vertex and Edge Approximate graph Miner, `VEAM`) is presented. `VEAM` addresses the approximate matching problem using both vertex and edge label sets during the mining process. The framework is tested in the context of graph-based image classification.

Considering the Object recognition task, in [30] an object is represented by means of `SIFT` features selected by an approach based on visual saliency.

Precisely the objects are modelled by a Class Specific Hyper-Graphs (CSHG) by exploiting Delaunay graphs and considering the SIFT keypoints as nodes.

In [19] another graph mining technique for object recognition is proposed. In this work an image is represented by an irregular pyramid. Each level of the structure is a RAG and the whole pyramid is built from bottom to top, where the base level is the entire image. Image regions are represented by different basic low-level descriptors to add context information and the structure is captured by employing a Frequent Approximate Subgraph (FAS).

In [20] a graph matching scheme that involves Visual Features (color, texture and shape) and Spatial Relations (VFSR) to detect similar objects is described. The goal is to show that the combination of visual and spatial features is a promising approach to improve the object recognition task. The spatial descriptors proposed are easy to build, store and manipulate, and can be employed to explicitly represent many possible spatial configurations between pairs of image regions, considering several basic orientation and topological relationships.

In [17] an object recognition approach based on hierarchical features is proposed(hereinafter RSW+Boosting) to capture local and structural information. This method employes a combination of decision trees to classify the objects.

In literature other approaches that capture local and spatial information without employing graph structures are presented. Among them, in [21] a temporal approach based on local features (hereinafter Sequential Patterns) is presented. The temporal information captures the spatial relations between local features. The problem of object recognition is seen as a sequential prediction task. A Discriminative Variable Memory Markov model, which captures multiple statistical sources features generating sequential patterns in a stochastic manner, is adopted.

3 Overview of our Framework

Our object recognition framework, called LLP + ARSRG*emb*, is composed by three different modules (the complete flowchart is shown in figure 1). In the first module each image is converted in an Attributed Relational SIFT-based Graph (ARSRG, [16]). This structure is able to capture local information preserving the spatial relationships between them (see Section 3.1). In the second module the set of ARSRGs is split in two subsets: training and prototypes. Each training ARSRG is embedded into an n-dimensional vector space (where n is the cardinality of prototypes set) by means of a novel graph embedding paradigm (see Section 3.2). Each component of this vector encodes the distance between the described structure and one of the prototypes. This distance is computed by an efficient graph matching algorithm proposed in [16]. Finally, a Logistic Label Propagation classifier (see Section 3.3) is trained on the n-dimensional vectors. Notice that, the computational complexity depends on the graph embedding of the

ARSRGs that is $O(N * M * K)$, where N are the number of ARSRG to encode into vector space, M are the ARSRG prototypes, and K is the time complexity of the ARSRG computation.

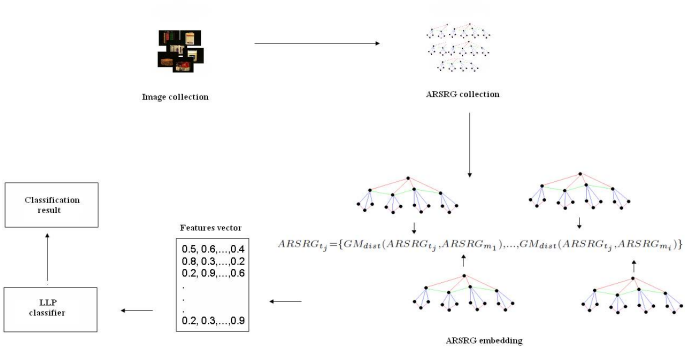

Fig. 1. Overview of the Object recognition framework

3.1 Graph Based Image Representation

In this section we describe the graph based image representation employed. This representation, called Attributed Relational SIFT-based Regions Graph (ARSRG), was proposed by Manzo et al. in [16]. The structure is composed by three different levels of nodes: the *Root node*, the *RAG Nodes*, and the *Leaf nodes*. The *Root node* represents the whole image and is linked with all the *RAG Nodes*[27] of the second level. *RAG Nodes* represent image regions, extracted by means of a segmentation technique, and encode adjacency relationships between them. At this level, adjacent regions in the image are represented by connected nodes. Finally, the *Leaf nodes* represent the set of SIFT [15] descriptors extracted from the image. Employing these descriptors invariance to the view-point, to the illumination, and to the scale is guaranteed. Precisely, a descriptor is associated to a region based on its spatial coordinates and the descriptors belonging to the same region are connected by edges. Figure 2 shows an example of ARSRG construction.

The choice of ARSRG for objects representation arises from an important property of the graph structure. This property concerns relations established among local features and structural information of the object encoded into the RAG configuration located at second level. It has been demonstrated that global configuration and local information of scene play a key role in the human recognition task. In this context, relations can be distinguished in: horizontal and vertical. Horizontal relations provide information about spatial closeness between image regions (level two) or SIFT features (level three). Vertical relations concern connections among image regions (level two) and SIFT features (level three). Using

this type of configuration invariance to changes, such as viewpoints, illuminations, scale, is ensured which are often the reason of poor performance in the task of objects recognition.

Compared to the approach proposed in [16], in which the ARSRG structures are adopted to calculate distances between images in order to solve a retrieval problem, we employ ARSRG structures to map image features, through the ARSRG embedding procedure described in the next section, in a space that can be easily managed during the classification stage.

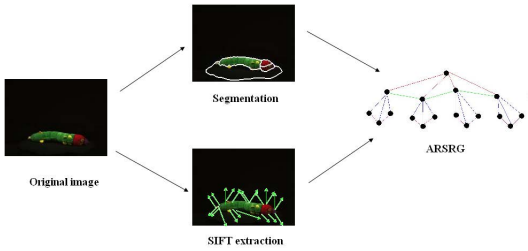

Fig. 2. An example of ARSRG construction

3.2 ARSRG Embedding

In literature many approaches have been proposed for dimensionality reduction. Among them, the most popular are Principal Component Analysis (PCA, [12]), Linear Discriminant Analysis (LDA, [2]), and Kernel variants of this techniques [2]. The main goal of these approaches is to derive lower dimensional representation from the original higher dimensional feature space preserving some properties of the data.

These techniques works on unstructured data. To overcome this limitation and handle structured data, such as graphs, we report the graph embedding approach, as proposed in [24] for non rigid scene recognition, with the purpose to provide a fixed-dimensional vector representation of an ARSRG structure.

Consider a labeled set of sample graphs $\boldsymbol{\mathcal{S}} = \{\boldsymbol{\mathcal{G}}_1, \ldots, \boldsymbol{\mathcal{G}}_n\}$ and a graph similarity measure $s(\boldsymbol{\mathcal{G}}_i, \boldsymbol{\mathcal{G}}_j)$, where $\boldsymbol{\mathcal{S}}$ can be any kind of graph set and $s(\boldsymbol{\mathcal{G}}_i, \boldsymbol{\mathcal{G}}_j)$ can be any kind of graph similarity measure. Moreover consider a set $\boldsymbol{\mathcal{P}} = \{\boldsymbol{\mathcal{P}}_1, \ldots, \boldsymbol{\mathcal{P}}_m\}$ of $m <= n$ prototypes extracted from $\boldsymbol{\mathcal{S}}$, and compute the similarities of a given input graph $\boldsymbol{\mathcal{G}}_j$ with each prototype $\boldsymbol{\mathcal{P}}_k \in \boldsymbol{\mathcal{P}}$. This leads to m similarities, $s_1 = s(\boldsymbol{\mathcal{G}}_j, \boldsymbol{\mathcal{P}}_1), \ldots, s_m = s(\boldsymbol{\mathcal{G}}_j, \boldsymbol{\mathcal{P}}_m)$, which can be represented in an m-dimensional vector $(s_1, \ldots, s_m)$. Employing this approach any graph can be transformed into a vector of real numbers. Precisely, consider a graph domain $\boldsymbol{\mathcal{G}}$, the training set of graphs $\boldsymbol{\mathcal{S}} = \{\boldsymbol{\mathcal{G}}_1, \ldots, \boldsymbol{\mathcal{G}}_n\} \subseteq \boldsymbol{\mathcal{G}}$, and a set of prototype graphs $\boldsymbol{\mathcal{P}} = \{\boldsymbol{\mathcal{P}}_1, \ldots, \boldsymbol{\mathcal{P}}_m\} \subseteq \boldsymbol{\mathcal{S}}$, the vector of mapping between $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{P}}$ is defined as follows:

$$\Phi_m^{\boldsymbol{\mathcal{P}}}(\boldsymbol{\mathcal{G}}_m) = (s(\boldsymbol{\mathcal{G}}_m, \boldsymbol{\mathcal{P}}_1), \ldots, s(\boldsymbol{\mathcal{G}}_m, \boldsymbol{\mathcal{P}}_m)) \tag{1}$$

where $s(\boldsymbol{\mathcal{G}}_m, \boldsymbol{\mathcal{P}}_i)$ is a graph similarity measure between graph $\boldsymbol{\mathcal{G}}_m$ and the ith prototype. This paradigm can be applied to ARSRG structures

obtaining a vector for each training ARSRG whose components encode the distance between the considered graph and all the ARSRG prototypes. Distance values are obtained through an iterative and efficient graph matching algorithm proposed in [16]. Precisely, this matching algorithm measures regions similarity among the ARSRG structures exploiting the topological relation information.

The matching phase is handled through a hierarchical exploration of ARSRG, that can be roughly divided in two steps: filtering of regions based on their size; subgraph matching performed by matching features belonging to single regions located at the third level of ARSRG.

The algorithm can be also seen as an image matching (retrieval) procedure, which works on two levels. The first level exploits global features, that are the regions extracted through a segmentation algorithm called JSEG [6], which performs a segmentation of color-texture regions in images through a first color quantization followed by a spatial segmentation; the second level explores local invariant region features. In this way, both local and structural image features are analyzed during the matching process.

The combination of graph embedding and graph based SIFT structures has already been proposed in the literature. In [9], graph based SIFT structures are embedded into a vector space according to the graph edit distance operations for generic object recognition application. Also, in [4] the graphs are mapped into a vector space by means of graph embedding, for representing human's shapes with purpose of action recognition. Differently from the aforementioned approaches, our algorithm is designed to solve more efficiently and effectively the graph embedding problem. Indeed our approach performs the matching phase employing the subgraphs representing image regions instead of the overall graphs representing the entire image, greatly reducing the time complexity. Moreover, since our approach considers local and structural information during the graph matching comparison the robustness to light, scale, and viewpoint changes is ensured (this does not happen for the aforementioned approaches that use edit operations). This is a key aspect that can strongly improve the object recognition performance.

3.3 Classification Phase

The final classification phase is managed by employing a semi-supervised learning technique called logistic label propagation (LLP, [13]) algorithm. This method employs the logistic function to classify input data, similarly to logistic regression. To deal with unlabeled samples as well as labeled ones, the logistic functions are learnt by using similarities between samples as proposed in [32]. Precisely, the learning problem is formulated as Gaussian random fields on graphs, where a field is described in terms of harmonic functions, and is efficiently solved using matrix methods or belief propagation. Note that, in this technique the logistic regression is effectively incorporated in terms of posterior probabilities.

4 Experimental Results

In this section we analyze the results achieved by our framework on three popular datasets. For each database we adopt the experimental settings proposed in well-known object recognition papers, especially the selection of graphs prototype crucial in the representation of objects-classes, thus to further assess our results. Moreover, we compare the achieved results with those obtained by a baseline approach that employs the same classifier used in our framework (Logistic Label Propagation classifier) applied on Bag of Words (`LLP` + `BoW`) to highlight the quality of our features. This section is organized as follows: in Section 4.1 the databases employed are summarized, while in Section 4.2 the results achieved are presented.

4.1 Datasets

The experiments have been performed on three datasets that differs in size, design, and topic. Precisely, we have employed the following databases:

1. The Columbia Image Database Library (COIL-100, [22]), which consists of 100 objects. Each object is represented by 72 colored images that show it under different rotation point of view.
2. The Amsterdam Library Of Images (ALOI, [8]) is a color image collection of 1000 small objects. We used the Object Viewpoint Collection. In contrast to COIL-100, where the objects are cropped to fill the full image, in ALOI the images contain the background and the objects in their original size.
3. The ETH-80 Image Set [14], which contains 80 objects from 8 categories and each object is represented by 41 different views, thus obtaining a total of 3280 images.

(a) (b) (c) (d) (e) (f)

Fig. 3. Example images from the COIL-100 dataset (first 2 images), ALOI dataset (second 2 images) and from the ETH-80 dataset (last 2 images)

4.2 Discussion

Table 1 summarizes the accuracy results of the proposed framework on ETH-80 database. In order to perform a direct comparison with the methods employed in [19], the same setup is adopted. Precisely, we took the same 6 categories

(*apples*, *cars*, *cows*, *cups*, *horses*, and *tomatoes*). For each category 4 objects are taken and for each object 10 different views are considered thus obtaining a total of 240 images. From the remaining images, 60 per category (15 views per object) are used as testing examples. We reported the results achieved by our baseline (LLP+BoW), and those obtained in [19] by employing the approaches proposed in [7] (gdFil), in [11] (APGM), and in [1] (VEAM). As can be seen in Table 1, our method outperforms the results obtained by the other approaches. These results confirm that our framework correctly deals with object view changes.

Table 1. Recognition accuracy on the ETH-80 database

Method	Accuracy
LLP+ARSRGemb	**89.26**%
LLP+BoW	58.83%
gdFil	47.59%
APGM	84.39%
VEAM	82.68%

Table 2 summarizes the results achieved by LLP + ARSRG*emb* on COIL-100 database. In order to perform a direct comparison with the methods employed in [19,20], the same setup is adopted. Precisely, we have randomly selected 25 objects and we have employed the 11% of the images as training set and the remaining ones as testing set. We have reported the results achieved by our baseline (LLP + BoW), and those obtained in [19,20] by employing their approach (VFSR) and the approaches proposed in [7] (gdFil), in [11] (APGM), in [1] (VEAM), in [29] (DTROD-AdaBoost), in [17] (RSW+Boosting), in [21] (Sequential Patterns), and in [23] (LAF). The results are presented in terms of accuracy and the best performance is highlighted in bold face. Our approach confirms its qualities also employing this database. Indeed our approach obtained the best overall accuracy.

Table 2. Recognition accuracy on the COIL-100 database

Method	Accuracy
LLP+ARSRGemb	**99.55**%
LLP+BoW	51.71%
gdFil	32.61%
VFSR	91.60%
APGM	99.11%
VEAM	99.44%
DTROD-AdaBoost	84.50%
RSW+Boosting	89.20%
Sequential Patterns	89.80%
LAF	99.40%

Table 3 summarizes the accuracy results obtained by LLP + ARSRG*emb* on the ALOI database. As can be see, the experiments have been performed by increasing the number of images thus to assess the robustness of the proposed framework. In order to perform a direct comparison with the methods employed in [28], the same setup is adopted; precisely, only the first 100 objects are employed. Color images have been converted to gray level and second image of each class was adopted for training and the remaining for testing. Two images of each class are considered, having a total of 200 images. Subsequently, at each iteration for each class one additional training image is attached. In Table 3 we have only shown the results by considering batch of 400 images since the intermediate results did not provide great differences. We have reported the results achieved by our baseline (LLP + BoW), and those obtained in [28] by employing some variants of Linear Discriminant Analysis (ILDAaPCA,batchLDA, ILDAonK, and ILDAonL). ILDAaPCA works first creating a PCA subspace by augmenting the k dimensional reconstructive subspace with additional $c-1$ vectors containing discriminative information. Those additional vectors are created from vectors that would be discarded when truncating the subspace to k-dimensions. In this way, the full discriminative information is included. Subsequently, the actual LDA representation from the obtained augmented subspace is built. batchLDA builds a new model in each update step using the same number of images as the incremental algorithms. ILDAonK updates a PCA basis truncated to the size $\hat{k} = k + c - 1$. The parameter k encodes the 80% of the energy, in term of fraction of the total variance, of the starting model constant during the experiment. ILDAonL updates the $(c-1)$-dimensional LDA basis directly and only discriminative information is used. These results show that our framework is able to obtain good results with a small amount of training set and that it is little affected by overfitting problems.

Table 3. Recognition accuracy on the ALOI database

Method	200	400	800	1200	1600	2000
LLP+ARSRGemb	**86.00%**	**90.00%**	**93.00%**	**96.00%**	**95.62%**	**96.00%**
LLP+BoW	49.60%	55.00%	50.42%	50.13%	49.81%	48.88%
batchLDA	51.00%	52.00%	62.00%	62.00%	70.00%	71.00%
ILDAaPCA	51.00%	42.00%	53.00%	48.00%	45.00%	50.00%
ILDAonK	42.00%	45.00%	53.00%	48.00%	45.00%	51.00%
ILDAonL	51.00%	52.00%	61.00%	61.00%	65.00%	69.00%

As can be noticed from the reported results, our framework is able to provide good overall performances for the Object recognition task, confirming its quality and its robustness. Moreover, this work confirms that capture local information preserving the spatial relationships between them can strongly improve the performance in the Object recognition field.

It is important to highlight that, thanks to the graph embedding paradigm, the main computational overhead only concerns the extraction of graph-based representation in the training stage, while the classification can be performed very quickly. This is particularly important since it was not possible to obtain results in a reasonable time when we tried to compare our framework with two state-of-the-art kernel graph approaches [3,26] on RAG structures.

5 Conclusions and Future Works

In this paper, we have proposed a framework to embed graph structures into vector spaces in order to improve object recognition. In this paper we have shown that representing images to efficiently capture global and local features improves the quality of the object recognition task and reduces the overfitting problems. Consequently, the achieved results compared to those obtained by well-known methodologies have shown that our approach is promising. Testing on larger and more complex databases such as Pascal VOC, Caltech 256, as well as the employment of different graph matching algorithms in order to improve the system performance are issues under exam.

References

1. Acosta-Mendoza, N., Gago-Alonso, A., Medina-Pagola, J.E.: Frequent approximate subgraphs as features for graph-based image classification. Knowledge-Based Systems **27**, 381–392 (2012)
2. Bishop, C.: Pattern Recognition and Machine Learning. Springer (2006)
3. Borgwardt, K.M., Kriegel, H.P.: Shortest-path kernels on graphs. In: Fifth IEEE International Conference on Data Mining, p. 8. IEEE (2005)
4. Borzeshi, E.Z., Piccardi, M., Xu, R.: A discriminative prototype selection approach for graph embedding in human action recognition. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp. 1295–1301. IEEE (2011)
5. Chatfield, K., Lempitsky, V., Vedaldi, A., Zisserman, A.: The devil is in the details: an evaluation of recent feature encoding methods. In: BMVC (2011)
6. Deng, Y., Manjunath, B.: Unsupervised segmentation of color-texture regions in images and video. IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(8), 800–810 (2001)
7. Gago-Alonso, A., Carrasco-Ochoa, J.A., Medina-Pagola, J.E., Martínez-Trinidad, J.F.: Full duplicate candidate pruning for frequent connected subgraph mining. Integrated Computer-Aided Engineering **17**(3), 211–225 (2010)
8. Geusebroek, J.M., Burghouts, G.J., Smeulders, A.W.: The amsterdam library of object images. International Journal of Computer Vision **61**(1), 103–112 (2005)
9. Hori, T., Takiguchi, T., Ariki, Y.: Generic object recognition by graph structural expression. In: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1021–1024. IEEE (2012)
10. Luo, J., Pronobis, A., Caputo, B., Jensfelt, P.: The kth-idol2 database. Technical Report CVAP304, Kungliga Tekniska Hoegskolan, CVAP/CAS (2006)
11. Jia, Y., Zhang, J., Huan, J.: An efficient graph-mining method for complicated and noisy data with real-world applications. Knowledge and Information Systems **28**(2), 423–447 (2011)
12. Jollife, I.T.: Principal Component Analysis. Springer Series in Statistics. Springer, New York (1986)
13. Kobayashi, T., Watanabe, K., Otsu, N.: Logistic label propagation. Pattern Recognition Letters **33**(5), 580–588 (2012)
14. Leibe, B., Schiele, B.: Analyzing appearance and contour based methods for object categorization. In: Proceedings. 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. II-409. IEEE (2003)

15. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60**(2), 91–110 (2004)
16. Manzo, M., Petrosino, A.: Attributed relational sift-based regions graph for art painting retrieval. In: Petrosino, A. (ed.) ICIAP 2013, Part I. LNCS, vol. 8156, pp. 833–842. Springer, Heidelberg (2013)
17. Marée, R., Geurts, P., Piater, J., Wehenkel, L.: Decision trees and random subwindows for object recognition. In: ICML Workshop on Machine Learning Techniques for Processing Multimedia Content (MLMM2005) (2005)
18. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(10), 1615–1630 (2005)
19. Morales-González, A., Acosta-Mendoza, N., Gago-Alonso, A., García-Reyes, E.B., Medina-Pagola, J.E.: A new proposal for graph-based image classification using frequent approximate subgraphs. Pattern Recognition **47**(1), 169–177 (2014)
20. Morales-González, A., García-Reyes, E.B.: Simple object recognition based on spatial relations and visual features represented using irregular pyramids. Multimedia Tools and Applications **63**(3), 875–897 (2013)
21. Morioka, N.: Learning object representations using sequential patterns. In: Wobcke, W., Zhang, M. (eds.) AI 2008. LNCS (LNAI), vol. 5360, pp. 551–561. Springer, Heidelberg (2008)
22. Nayar, S.K., Nene, S.A., Murase, H.: Columbia object image library (coil 100). Department of Comp. Science, Columbia University, Tech. Rep. CUCS-006-96 (1996)
23. Obdrzalek, S., Matas, J.: Object recognition using local affine frames on distinguished regions. In: BMVC, vol. 2, pp. 13–122 (2002)
24. Pellino, S., Petrosino, A.: Bag of graph words for scene recognition. Pattern Recognition Letters p. submitted (2014)
25. Rozza, A., Manzo, M., Petrosino, A.: A novel graph-based fisher kernel method for semi-supervised learning. In: Submitted to ICPR 2014 (2014)
26. Shervashidze, N., Schweitzer, P., Van Leeuwen, E.J., Mehlhorn, K., Borgwardt, K.M.: Weisfeiler-lehman graph kernels. The Journal of Machine Learning Research **12**, 2539–2561 (2011)
27. Tremeau, A., Colantoni, P.: Regions adjacency graph applied to color image segmentation. IEEE Transactions on Image Processing **9**(4), 735–744 (2000)
28. Uray, M., Skocaj, D., Roth, P.M., Bischof, H., Leonardis, A.: Incremental lda learning by combining reconstructive and discriminative approaches. In: BMVC, pp. 1–10 (2007)
29. Wang, Y., Gong, S.: Tensor discriminant analysis for view-based object recognition. In: 18th International Conference on Pattern Recognition, ICPR 2006, vol. 3, pp. 33–36. IEEE (2006)
30. Xia, S., Hancock, E.: 3d object recognition using hyper-graphs and ranked local invariant features. Structural, Syntactic, and Statistical Pattern Recognition, 117–126 (2008)
31. Xiao, J., Hays J., Ehinger, K., Oliva, A., Torralba, A.: Sun database: Large-scale scene 488 recognition from abbey to zoo. In: IEEE Conference on Computer Vision and Pattern 489 Recognition (CVPR), pp. 3485–3492 (2010)
32. Zhu, X., Ghahramani, Z., Lafferty, J., et al.: Semi-supervised learning using gaussian fields and harmonic functions. In: ICML, vol. 3, pp. 912–919 (2003)