Attributed Relational SIFT-Based Regions Graph for Art Painting Retrieval

Mario Manzo and Alfredo Petrosino

Department of Science and Technology
University of Naples "Parthenope",
Isola C4, Centro Direzionale, Napoli, Italy
{mario.manzo,alfredo.petrosino}@uniparthenope.it

Abstract. Recently, image retrieval and analysis algorithms have been extensively applied to art related domains. In this field, state-of-the-art approaches mainly focus on feature extraction with the aim of improving reliability of authentication, classification and retrieval of art paintings. In this paper we propose an effective modeling, based on a graph structure, and a retrieval strategy, based on a graph matching algorithm, for art paintings. The proposed approach has been tested on different datasets with high quality results allowing an user to run effective content-based queries on painting records.

Keywords: CBIR systems, graph matching, graph based image representation, local invariant features extraction.

1 Introduction

During last decade, computer graphics and vision experts have focused their attention on the problem of cultural heritage preservation. In this field, effective techniques have been proposed concerning classification [11] and retrieval [20]. In [4] a graph-based method is described for automatic annotation and retrieval of digital art print images. This method has been proved to be particularly useful for art historians to annotate database of digital art print images. In [8] a colorimetric visualization method is proposed based on a spatial organization of colors within the painting. The effectiveness of the method is evaluated on Italian Renaissance images. Other approaches exploits Local Invariant Features Extraction (LIFE) methods [18] for image representation and similarity measurements. Authors in [11] present a novel approach for painting classification based on image segmentation and SIFT [16]/SURF [2] features extraction. In [20] a system for retrieving information about paintings using mobile devices is presented. An augmented reality system based on SIFT[16] is described in [24] to retrieve information about artist and historical context of paintings.

In this paper we propose a novel graph-based image representation along with a graph matching algorithm to effectively tackle the art painting retrieval task. A segmented digital image can be seen as a set of regions, each carrying two types of information: local visual information (color, shape or texture)

A. Petrosino (Ed.): ICIAP 2013, Part I, LNCS 8156, pp. 833–842, 2013.

and spatial global information (topological configuration of regions located in a neighborhood). Indeed, relations between local and global information play a key role in human recognition task [13]. Many approaches [4] represent images using a graph structure considering, in this way, the image matching problem as a graph matching problem. In this context, the aim of our paper is threefold: first, a graph structure for image representation called Attributed Relational SIFT-based Regions Graph (ARSRG) is introduced to reduce the gap between local and global features; second a graph matching algorithm is presented to measure regions similarity exploiting information about topological relations; last, the LIFE method is applied in order to extract stable descriptors starting from a given set of image features.

The paper is organized as follows. Section 2 describes the graph based image representation, while Section 3 describes the graph matching algorithm. Relevant results are discussed in Section 4 and conclusions are drawn in Section 5.

2 Graph Based Image Representation

In this section we introduce a novel graph based image representation, composed by two main steps: features extraction and graph construction.

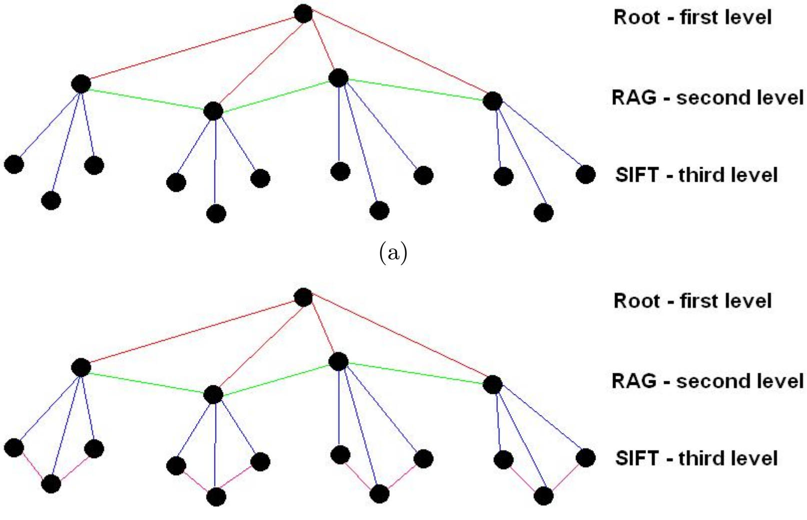
The first step consists in the extraction of the regions of interest (ROIs) from an image, by means of a segmentation technique, and the construction of a *Region Adjacency Graph (RAG)*[23] to encode spatial relations between extracted regions.

The second step consists of the construction of a graph, named by us **Attributed Relational SIFT-based Regions Graph (ARSRG)**, composed by three levels: *Root node*, *RAG Nodes* and *Leaf nodes*. The *Root node* represents the whole image and is linked to all the *RAG Nodes* at second level. *RAG Nodes* encode adjacency relationships between different image regions. Thus, adjacent regions in the image are represented by connected nodes. Finally, *Leaf nodes* represent the set of SIFT descriptors extracted from the image, in order to tackle invariance to view-point, illumination and scale.

Two types of configurations are provided at this level: *Region based* and *Region graph based* (Figure 1). In *Region based* a keypoint is associated to a region based on its spatial coordinates, whereas, *Region graph based* contains keypoints belonging to the same region connected by edges, which encode spatial adjacency. ARSRG can be defined by structures based on two different *Leaf nodes* configurations.

Definition 1. *An* $\boldsymbol{ARSRG}_{1^{st}}$ *(first leaf nodes configuration),* G *is defined as a tuple* $G = (V_{regions}, E_{regions}, VF_{SIFT}, E_{regions-SIFT})$*, where:*

- $V_{regions}$*, the set of regions-nodes.*
- $E_{regions} \subseteq V_{regions} \times V_{regions}$*, the set of undirected edges, where* $e \in E_{regions}$ *and* $e = (v_i, v_j)$ *is an edge between nodes* $v_i, v_j \in V_{regions}$*.*
- VF_{SIFT}*, the set of SIFT-nodes.*



(a)

(b)

Fig. 1. *Region based* (a) and *Region graph based* (b) Configurations

- $E_{regions-SIFT} \subseteq V_{regions} \times VF_{SIFT}$, *the set of directed edges, where* $e \in E_{regions-SIFT}$ *and* $e = (v_i, vf_j)$ *is an edge between source node* $v_i \in V_{regions}$ *and destination node* $vf_j \in VF_{SIFT}$.

Definition 2. *An* $\boldsymbol{ARSRG}_{2^{nd}}$ *(second leaf nodes configuration), G is defined as a tuple* $G = (V_{regions}, E_{regions}, VF_{SIFT}, E_{regions-SIFT}, E_{SIFT})$, *where:*

- $V_{regions}$, *the set of regions-nodes.*
- $E_{regions} \subseteq V_{regions} \times V_{regions}$, *the set of undirected edges, where* $e \in E_{regions}$ *and* $e = (v_i, v_j)$ *is an edge between nodes* $v_i, v_j \in V_{regions}$
- VF_{SIFT}, *the set of SIFT-nodes.*
- $E_{regions-SIFT} \subseteq V_{regions} \times VF_{SIFT}$, *the set of directed edges, where* $e \in E_{regions-SIFT}$ *and* $e = (v_i, vf_j)$ *is an edge between source node* $v_i \in V_{regions}$ *and destination node* $vf_j \in VF_{SIFT}$.
- $E_{SIFT} \subseteq VF_{SIFT} \times VF_{SIFT}$, *the set of undirected edges, where* $e \in E_{SIFT}$ *and* $e = (vf_i, vf_j)$ *is an edge between nodes* $vf_i, vf_j \in V_{SIFT}$

The ARSRG structure has a set of properties arising from two building blocks: features extraction and matching. The first building block includes relations among local features and structural information of image 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 [13]. Relations can be distinguished in: horizontal and vertical. Horizontal relations provide information about spatial closeness between ROIs (level two) or SIFT features (level three). Vertical relations concern connections among ROIs (level two) and SIFT features (level three). The second building block includes an algorithm for optimal matching and false positive reduction, to refine results. 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 of ARSRG.

3 Graph Matching

Given two ARSRGs, the goal is to find best matches among their nodes and to determine a mapping set M containing associated nodes between the two structures. This is done by iterative exploration of best possible nodes mapping and selecting the best pairs at each iteration, adopting two approaches to measure dissimilarity among node pairs: ratio test[16] and graph matching[21].

The first step of the algorithm is the construction of a $n \times m$ matrix, called $Dist\_matrix$, where n and m are the numbers of regions-nodes at the second level of the two ARSRGs respectively. The matrix contains the distances between each node of the first ARSRG and all the nodes of the second ARSRG. In order to find the most promising mapping, a second matrix B, of dimension $n \times m$, stores the mapping corresponding to the minimum value of rows in $Dist\_matrix$. For each possible nodes mapping extracted from B, the algorithm computes matches generated by SIFT descriptors associated to the nodes. Nodes pairs that present a number of matches greater than a given threshold are saved.

Next, the algorithm analyzes the second-smallest elements at each row of matrix $Dist\_matrix$ extracting, from B, the correspondences that contain at least one node-to-node matching and so on, until it reaches the final iteration.

3.1 Regions Matching with Ratio Test

Different approaches can be employed to find the best match for each region. For instance, given two regions with associated SIFT keypoints, the naive approach consists in searching for the best candidate match for each keypoint in the first region by identifying its nearest neighbor in the second region, using a global threshold. This approach produces many false matches, i.e. many keypoints do not match correctly due to the global threshold. Therefore, a different measure is adopted comparing the closest to the second-closest neighbor of each keypoint [16].

3.2 Regions Matching with Graph Matching

Differently from the previous solution, the problem of regions comparison can be reformulated in terms of graph matching [21], with the goal of improving the quality of the matches. We consider SIFT features organized in the form of SIFT Nearest Neighbor Graph (SNNG) according to the following definition:

Definition 3. *A $SNNG = (VF_{SIFT}, E_{SIFT})$ is defined as*

- VF_{SIFT}*: the set of nodes associated to SIFT keypoints*
- E_{SIFT}*: the set of edges*

An edge $e = (v_i, v_p)$ exists, for $v_i, v_p \in VF_{SIFT}$, if $dist(v_i, v_p) < \tau$, where $dist(v_i, v_p)$ is the Euclidean distance, τ is a threshold value and p stems from 1 *to k, k being the size of VF_{SIFT}.*

SNNG represents SIFT keypoints belonging to image region located at the third level of ARSRG structure according to Definition 2. Matches among SNNGs are described through a matrix S that defines an injective mapping between two SNNGs: $SNNG_1 = (VF_{SIFT1}, E_{SIFT1})$ and $SNNG_2 = (VF_{SIFT2}, E_{SIFT2})$. In particular, if an element $s_{ij} \in S$ is assigned to 1 then the node $v_i \in VF_{SIFT1}$ matches with node $v_j \in VF_{SIFT2}$, otherwise 0. In this context, the goal of algorithm is to initially estimate best matrix S, starting from the initial guess $S^{(1)}$ through the space of matching configurations. We use a combined measure of structural consistency and similarity called W, to compare SNNGs during the matching. Given two nodes $v_a \in VF_{SIFT1}$ and $v_\alpha \in VF_{SIFT2}$, we define

$$W_{a\alpha} = Q_{a\alpha} R_{a\alpha} \tag{1}$$

where

$$Q_{a\alpha} = exp\left[\mu \sum_{b \in V_1} \sum_{\beta \in V_2} D_{ab} M_{\alpha\beta} s_{b\beta}\right] \qquad and \qquad R_{a\alpha} = \frac{1}{dist(z_a^1, z_\alpha^2)} \tag{2}$$

$Q_{a\alpha}$ is the structural consistency coefficient, D and M are the adjacency matrices of G_1 and G_2, $s_{b\beta}$ is an element of matrix S and $\mu > 0$ is a control parameter. $R_{a\alpha}$ is a similarity nodes matching function, where $dist(z_a^1, z_\alpha^2)$ is the Euclidean distance between SIFT descriptors z_a^1 and z_α^2 corresponding to nodes v_a and v_α.
Moreover, in order to describe the matching node-by-node between two SNNGs, an additional matrix Ω is adopted

$$\Omega = \begin{bmatrix} W_{11} & \cdots & W_{1m} \\ \vdots & W_{a\alpha} & \vdots \\ W_{n1} & \cdots & W_{nm} \end{bmatrix} \tag{3}$$

A cleaning heuristic approach to extract best matches is applied on Ω with the purpose of building matrix S. The iterative procedure is composed by three steps:

1. at the first step, $W_{a,k} = max(W_{a,\alpha})$ is selected at each row a of Ω, $\alpha = 1, ..., m$, such that $W_{a,k}/W_{a,k2} > \frac{1}{\rho}$, where $W_{a,k2}$ is the second greatest element in the a-th row of Ω;

2. the second step finds the maximum element $W_{a,\alpha} \in \Omega$ and activates the corresponding match $s_{a\alpha} \in S$;
3. at the third step, the rows and columns of Ω containing $W_{a,\alpha}$ are sets to zero.

The three steps are repeated until Ω does not contain any other element to analyze, i.e. $W_{ij} = 0$, $\forall i, j$ $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

4 Experimental Results

The proposed approach has been tested on three datasets and compared with other LIFE methods, graph matching algorithms and CBIR system reported in the literature. The first dataset, described in [11], is composed by two sets of images obtained from Olga's gallery[1] and Travel Webshots[2]. The second dataset, described in[10], is composed by painting photos taken from the Cantor Arts Center[3]. The third dataset, described in [20], is composed by 1002 images. Figure 2 shows some examples.

4.1 LIFE Methods Comparison

A first evaluation is performed for dataset used in [11] and through comparisons with LIFE methods. Results are reported in terms of Mean Reciprocal Rank (MRR). As in [11,16], a tuning procedure is applied to ρ parameter that controls tolerance of false matches both in graph matching and ratio test. We used the values of ρ as suggested in [11] and [16]. In particular, ρ values of 0.6 and 0.7 are used in [11] and values greater than 0.8 are rejected as in [16].

Table 1. Quantitative comparison using *MRR* measure among SIFT[16], SURF[2], ORB[19], FREAK[1], BRIEF[3] and ARSRG matching on dataset in[11]

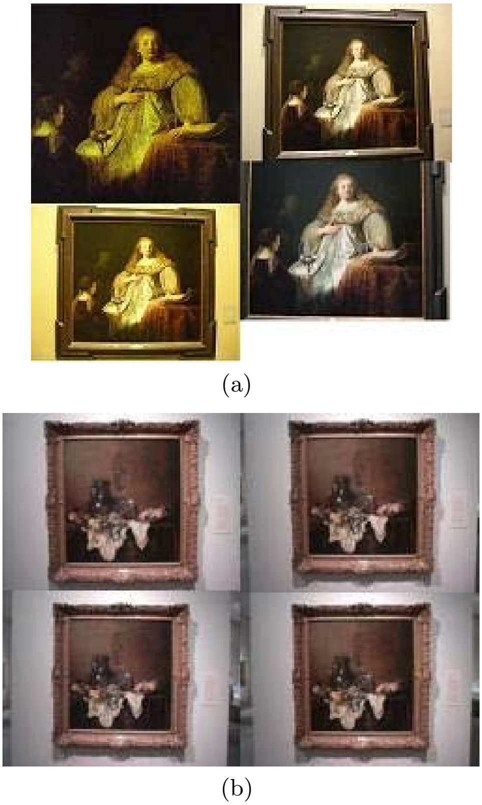
ρ	$SIFT$[16]	$SURF$[2]	ORB[19]	$FREAK$[1]	$BRIEF$[3]	$ARSRG_{1st}$	$ARSRG_{2nd}$
0.6	0.7485	0.8400	0.6500	0.3558	0.4300	0.6700	0.6750
0.7	0.7051	0.6800	0.6116	0.3360	0.3995	0.7133	0.7500
0.8	0.6963	0.5997	0.5651	0.2645	0.4227	0.6115	0.8000

Table 1 shows that graph based approach provides best performance. ρ values of 0.7 and 0.8 give optimal results for ARSRG matching. Graph based image representation clearly captures the topological relationships among features and acts as a filter over the complete set of SIFT features extracted from the image. Indeed, the comparison was performed among descriptors belonging to regions

[1] http://www.abcgallery.com/index.html

[2] http://travel.webshots.com

[3] http://museum.stanford.edu/



(a)

(b)

Fig. 2. Some examples of art painting images

instead of entire image as proposed in standard approaches. In this way, many false matches are discarded and effectiveness is greatly improved.

A second test has been performed on the dataset adopted in [10], computing performance in terms of Precision and Recall. Values of ρ parameter are the same as in the previous test.

Table 2. Quantitative comparison, using *Recall* measure, among SIFT[16], SURF[2], ORB[19], FREAK[1], BRIEF[3] and ARSRG matching on dataset in[10]

ρ	$SIFT$[16]	$SURF$[2]	ORB[19]	$FREAK$[1]	$BRIEF$[3]	$ARSRG_{1st}$	$ARSRG_{2nd}$
0.6	1.0	0.8666	0.8000	0.7333	0.7666	0.7333	0.7333
0.7	1.0	0.9000	0.8666	0.7333	0.8666	0.7666	0.7333
0.8	1.0	1.0	1.0	0.8333	1.0000	0.8000	0.8000

Table 2 shows that SIFT based approach performs better in terms of Recall. In case of ρ equal to 0.8, our approach yields comparable results.

Table 3. Quantitative comparison using *Precision* measure, among SIFT[16], SURF[2], ORB[19], FREAK[1], BRIEF[3] and ARSRG matching on dataset in[10]

ρ	$SIFT$[16]	$SURF$[2]	ORB[19]	$FREAK$[1]	$BRIEF$[3]	$ARSRG_{1^{st}}$	$ARSRG_{2^{nd}}$
0.6	0.0674	0.0820	0.2051	0.05584	0.10689	1.0	1.0
0.7	0.0401	0.0441	0.0742	0.04671	0.05664	0.6571	1.0
0.8	0.0312	0.0338	0.0348	0.04072	0.03452	0.1428	0.6666

In contrast, Table 3 shows that our approach, clearly outperforming the other approaches in terms of Precision, proves to be very effective for image retrieval problem. The best results by graph matching algorithm for Precision are provided with ρ equal to 0.6 and 0.7. These results are due to the use of image structural representation. Indeed, graph nodes, representing different image regions, provide a partitioning rule applied on entire set of SIFT. In this way, the subsets obtained are considered separately during matching step. This strategy removes most of false matches that normally belongs to accepted matches. As a consequence, several images are discarded as candidates for final ranking.

4.2 Graph Matching Algorithms Comparison

This section describes performance comparison with graph SIFT-based matching algorithms. Experiments are performed on datasets presented in [11,20] and are evaluated through MRR measure. Results are reported in tables 4 and 5.

Table 4. Quantitative comparison, using *MRR* measure, among HGM[14], RRWGM[15], TM[9] algorithms and ARSRG matching on dataset in[11]

HGM[14]	$RRWGM$[15]	TM[9]	$ARSRG_{1^{st}}$	$ARSRG_{2^{nd}}$
0.2600	0.1322	0.1348	0.6115	0.8000

Table 5. Quantitative comparison, using *MRR* measure, among HGM[14], RRWGM[15], TM[9] algorithms and ARSRG matching on dataset in[20]

HGM[14]	$RRWGM$[15]	TM[9]	$ARSRG_{1^{st}}$	$ARSRG_{2^{nd}}$
0.1000	0.0545	0.0545	0.20961	0.39803

In particular, Tables 4 and 5 show comparison, in terms of MMR values, with HGM [14], RRWGM [15], TM [9] algorithms. Also in this case, ARSRG leads to better results compared to those obtained by the other graph SIFT-based matching algorithms. Similarly in this case, the region matching approach, by providing local information about spatial distribution of the features, leads to false matches removal and hence improves final results.

4.3 CBIR System Comparison

This section describe the performance comparison with Lucene Image Retrieval (LIRe) [17] system. Experiments are performed on dataset presented in [11], considering different features implemented in LIRe, and evaluated through MRR measure. Results are reported in table 6.

Table 6. Quantitative comparison using MRR measure, among some features available in (LIRe)[17] system and ARSRG matching on dataset in[11]

$MPEG7$[5]	$Tamura$[22]	$CEDD$[6]	$FCTH$[7]	ACC[12]	$ARSRG_{1^{st}}$	$ARSRG_{2^{nd}}$
0.2645	0.1885	0.2329	0.1924	0.1879	0.7133	0.7500

From the reported results, it is clear that LIRe system is not very suitable for art paint retrieval, due to its low performing features, which results in wrong discrimination of relevant and irrelevant images. Consequently, the achieved ranking contains inadequate results, with respect to user's request, which affects heavily its final performance. In contrast, results obtained by ARSRG algorithm, demonstrates once more that the proposed approach is very effective for this application.

5 Conclusion Remarks

In this paper a novel way to capture visual and structural information from digital art paintings has been proposed. The resulting ARSRG structure has proved to be a valid alternative to standard techniques which use color, shape and texture to describe image content. Robustness and effectiveness of the proposed graph matching algorithm have been extensively tested on different public data repositories for the art painting retrieval task. The proposed approach is robust to changes in scale and lighting conditions, and allows to effectively retrieve objects based on the user preferences.

References

1. Alahi, A., Ortiz, R., Vandergheynst, P.: FREAK: Fast Retina Keypoint. In: CVPR, pp. 510–517 (2012)
2. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
3. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: Binary Robust Independent Elementary Features. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 778–792. Springer, Heidelberg (2010)
4. Carneiro, G.: Graph-based methods for the automatic annotation and retrieval of art prints. In: ICMR, p. 32 (2011)

5. Chang, S.F., Sikora, T., Puri, A.: Overview of the mpeg-7 standard. Circuits and Systems for Video Technology 11(6), 688–695 (2001)
6. Chatzichristofis, S.A., Boutalis, Y.S.: CEDD: Color and Edge Directivity Descriptor. A Compact Descriptor for Image Indexing and Retrieval. In: Gasteratos, A., Vincze, M., Tsotsos, J.K. (eds.) ICVS 2008. LNCS, vol. 5008, pp. 312–322. Springer, Heidelberg (2008)
7. Chatzichristofis, S.A., Boutalis, Y.S.: FCTH: Fuzzy Color And Texture Histogram A Low Level Feature For Accurate Image Retrieval. In: 9th International Workshop on Image Analysis for Multimedia Interactive Services, pp. 191–196 (2008)
8. Colantoni, P., Jean-Baptiste, T., Ruven, P.: Graph-based 3d visualization of color content in paintings. In: VAST, pp. 25–30 (2010)
9. Duchenne, O., Bach, F., Kweon, I.S., Ponce, J.: A tensor-based algorithm for high-order graph matching. PAMI 33(12), 2383–2395 (2011)
10. Etezadi-Amoli, M., Chang, C., Hewlett, M.: A day at the museum (2009)
11. Haladová, Z., Šikudová, E.: Limitations of the SIFT/SURF based methods in the classifications of fine art paintings. Computer Graphics and Geometry 12(1), 40–50 (2010)
12. Huang, J., Kumar, S.R., Mitra, M., Zhu, W.J., Zabih, R.: Image indexing using color correlograms. In: CVPR, pp. 762–768 (1997)
13. Koffka, K.: Principles of Gestalt Psychology. Harcourt, New York (1935)
14. Lee, J., Cho, M., Lee, K.M.: Hyper-graph matching via reweighted random walks. In: CVPR, pp. 1633–1640 (2011)
15. Cho, M., Lee, J., Lee, K.M.: Reweighted random walks for graph matching. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part V. LNCS, vol. 6315, pp. 492–505. Springer, Heidelberg (2010)
16. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
17. Lux, M., Chatzichristofis, S.A.: Lire: Lucene image retrieval: an extensible Java CBIR library. In: 16th ACM Multimedia, pp. 1085–1088 (2008)
18. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. PAMI 27(10), 1615–1630 (2005)
19. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.R.: ORB: An efficient alternative to SIFT or SURF. In: ICCV, pp. 2564–2571 (2011)
20. Ruf, B., Kokiopoulou, E., Detyniecki, M.: Mobile museum guide based on fast SIFT recognition. In: Detyniecki, M., Leiner, U., Nürnberger, A. (eds.) AMR 2010. LNCS, vol. 5811, pp. 170–183. Springer, Heidelberg (2010)
21. Sanroma, G., Alquézar Mancho, R., Serratosa, I., Casanelles, F.: Graph matching using SIFT descriptors - An application to pose recovery of a mobile robot. In: 5th International Conference on Computer Vision Theory and Applications, pp. 249–254 (2010)
22. Tamura, H., Mori, S., Yamawak, T.: Textural features corresponding to visual perception. Systems, Man, and Cybernetics 8(6), 460–472 (1978)
23. Tremeau, A., Colantoni, P.: Regions adjacency graph applied to color image segmentation. Trans. on Image Processing 9(4), 735–744 (2000)
24. You, S., Neumann, U.: Mobile augmented reality for enhancing e-learning and e-business. In: International Conference on Internet Technology and Applications, pp. 1–4 (2010)