Accepted version of paper published in International Conference on Multimedia Retrieval (ICMR), 2016.
DOI: http://dx.doi.org/10.1145/2911996.2912060

Rank Diffusion for Context-Based Image Retrieval

Daniel Carlos Guimarães Pedronette
Dept. of Statistic, Applied Math. and Computing
Universidade Estadual Paulista (UNESP)
Rio Claro, Brazil
daniel@rc.unesp.br

Ricardo da S. Torres
Recod Lab - Institute of Computing
University of Campinas (UNICAMP)
Campinas, Brazil
rtorres@ic.unicamp.br

ABSTRACT

This paper presents an efficient diffusion-based re-ranking approach. The proposed method propagates contextual information defined in terms of top-ranked objects of ranked lists in a diffusion process. That makes the method suitable for large scale real-world collections. Experiments were conducted considering public image collections, several descriptors, and comparisons with state-of-the-art methods. Experimental results demonstrate that the proposed method provides high effectiveness gains with low computational costs.

Keywords

content-based image retrieval; unsupervised distance learning; rank diffusion process

1. INTRODUCTION

A relevant change can be observed in the multimedia content generation, since common users are not long mere consumers and have become active producers. As a result, huge amounts of visual content have been accumulated daily, generated from a large variety of digital sources. The Content-Based Image Retrieval (CBIR) systems are considered a promising solution in this scenario, supporting searches capable of taking into account the visual properties of digital images.

The development of CBIR systems has been mainly supported by the creation of various visual features (based on shape, color, and texture properties) and different distance measures. However, more recently, research initiatives have focused on other stages of the retrieval process, which are not directly related to low-level feature extraction procedures. In several computer vision and image retrieval applications, capturing and exploiting the intrinsic manifold structure becomes a central problem for different vision, learning, and retrieval tasks [14].

Diverse methods also have been proposed in order to improve the effectiveness of distance measures in image retrieval tasks, specially based on diffusion approaches [14, 15, 39, 40]. In general, these post-processing methods aims at replacing pairwise distances by more global affinity measures capable of considering the dataset manifold [40]. Although indispensable for improving the effectiveness of retrieval, the wide use of post-processing methods on large-scale real-world applications also depends on efficiency and scalability aspects [24]. More recently, due to the high computational costs associated with diffusion-based approaches, other methods have emerged. Mainly based on rank analysis, such contextual rank measures [2, 26] can be efficiently computed.

In this paper, we propose a novel hybrid approach, named Rank Diffusion, which is based on a diffusion process defined in terms of ranking information. This method establishes a relationship between diffusion approaches and contextual rank measures in the sense it spreads ranking information, taking into account the dataset structure. A relevant contribution of the proposed approach is given in terms of efficiency aspects. Despite the use of a diffusion strategy, since only rank information is considered, low computational efforts are required.

An experimental evaluation was conducted, considering public datasets and several image descriptors, including global, local, and convolution-neural-network-based descriptors. Experiments were conducted on different retrieval tasks. The proposed method achieved significant effectiveness gains, yielding better results in terms of effectiveness performance than state-of-the-art approaches. For example, an accuracy score of 100% and a N-S score of 3.94 were achieved on the well-known MPEG-7 [17] and UKBench [22] datasets.

2. RELATED WORK

In various multimedia and learning applications, objects are often modeled as high dimensional points in an Euclidean space. For retrieval or classification purposes, the distance between two objects is computed often considering the Euclidean distance. However, once pairwise distance measures define relationships only between pairs of images, the global structure of the dataset and the context wherein the query is computed are ignored. Therefore, how to capture and exploit the intrinsic manifold structure becomes a central problem in the vision and retrieval applications [14]. In this scenario, many approaches have been proposed [6, 9, 12, 14, 26, 28, 38, 39] to improve the effectiveness of image retrieval tasks. Such methods take into account the dataset manifold and the global relationships among images, without the need of any user intervention. Some of the most important diffusion- and rank-based methods.

ICMR'16, June 06-09, 2016, New York, NY, USA
© 2016 ACM. ISBN ...$15.00
DOI:

Diffusion-based approaches [7] rely on the definition of a global measure, which describes the relationship between pairs of points in terms of their connectivity. In general, diffusion processes consider as input a pairwise affinity matrix W, which can be interpreted as a graph that encodes the relationships among objects [9]. Let $G = (V, E)$ be a graph, such as the nodes $v_i \in V$ are associated with dataset objects and edges $e_{ij} \in E$ indicate the existence of relationships among them. Edge weights, in turn, are defined by the affinity values w_{ij}. The matrix W is often computed by applying a Gaussian kernel to a distance matrix computed by an image descriptor [14]. Giving the edge weights [9] defined by the matrix W, the diffusion processes spread the affinities through the graph. In general, a walk in the graph occurs more likely through the edges with larger weights.

Several methods based on diffusion approaches have been proposed [9, 14, 38, 39]. The diffusion-based algorithms have been achieving significant improvements on retrieval performance although they are very expensive to compute, specially when the size of datasets becomes larger [2, 27].

Recently, various contextual rank-based approaches [2, 6, 26, 27, 29] have yielded very significant gains on retrieval effectiveness. Additionally, since the most relevant information of rankings is found at top positions, the rank-based approaches can significantly reduce the computational efforts required by exploiting indexing structures [24]. Therefore, other important requirements, such as efficiency and scalability [2, 24], are met.

In this paper, a novel hybrid re-ranking approach, which is based on a diffusion process defined in terms of ranking information, is presented. An advantage of the proposed method when compared with other approaches rely on its efficiency. As the diffusion process is based o rank information, only the top-ranked images can be considered.

3. FORMAL PROBLEM DEFINITION

This section formally defines the image retrieval and ranking model. Let $\mathcal{C}$=$\{img_1, img_2, \ldots, img_n\}$ be an image collection. The notation $\rho(i, j)$ denotes the distance between two images img_i and img_j according to a given descriptor. Let img_q be a query image. A ranked list τ_q can be computed in response to img_q based on the distance function ρ. The ranked list σ_q=(img_1, img_2, $\ldots$, img_n) can be defined as a permutation of the collection $\mathcal{C}$. A permutation σ_q is a bijection from the set $\mathcal{C}$ onto the set $[N] = \{1, 2, \ldots, n\}$. For a permutation σ_q, we interpret $\sigma_q(i)$ as the position (or rank) of image img_i in the ranked list σ_q.

The top positions of ranked lists are expected to contain the most similar images to the query image. Additionally, σ_q is expensive to compute, specially when n is high. Therefore, the computed ranked lists can consider only a sub-set of the images. Let τ_q be a ranked list that contains only the L most similar images to img_q, where $L \ll n$. Formally, let $\mathcal{C}_L$ be a sub-set of of the collection $\mathcal{C}$, such that $\mathcal{C}_L \subset \mathcal{C}$ and $|\mathcal{C}_L| = L$. The ranked list τ_q can be defined as a bijection from the set $\mathcal{C}_L$ onto the set $[N] = \{1, 2, \ldots, L\}$. Every image $img_i \in \mathcal{C}$ can be taken as a query image img_q. A set of ranked lists $\mathcal{T}$ = $\{\tau_1, \tau_2, \ldots, \tau_n\}$ can also be obtained, with a ranked list for each image in the collection $\mathcal{C}$.

Our goal is to exploit the similarity information encoded in the set $\mathcal{T}$, with the aim of computing a more effective set $\mathcal{T}_r$. In fact, a more effective distance function ρ_r is defined, giving rise to a new set of ranked lists. Additionally, the fusion problem is also considered, in which different sets of ranked lists $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_d\}$ are taken as input aiming at computing a more effective set $\mathcal{T}_r$.

4. RANK DIFFUSION PROCESS

This section discusses and defines the proposed rank diffusion process, presenting each step involved in the method, until the computation of the rank-diffusion distance.

4.1 Rank Similarity Matrix

Many diffusion-based algorithms use the distance information for defining a similarity matrix W. A Gaussian kernel is often considered, such that $w_{ij} = exp(\frac{-\rho^2(i,j)}{2\sigma^2})$, where σ is a parameter to be defined. Therefore, some strategies are required to define a suitable value for the parameter, also considering that the distance distribution may vary among different descriptors.

In this work, a rank similarity matrix W is proposed based only on rank information. The rank modeling of similarity information allows an uniform representation, independent of distance scores. The similarity score w_{ij} varies linearly according to the position of img_j in the ranked list τ_i. Additionally, the score considers only a neighborhood set, which is defined by the size of ranked lists.

Let m denote the size of ranked lists and, therefore the neighborhood considered. Let $\mathcal{N}(i, m)$ be the neighborhood set, which is formally defined as follows:

$$\mathcal{N}(i, m) = \{\mathcal{R} \subseteq \mathcal{C}, |\mathcal{R}| = m \wedge \forall x \in \mathcal{R}, y \in \mathcal{C} - \mathcal{R} : \rho(i, x) \leqslant \rho(i, y)\}. \tag{1}$$

The similarity rank matrix W_m is defined as:

$$w_{m_{ij}} = \begin{cases} m - \tau_i(j) + 1 & if\ img_j \in \mathcal{N}(i, m) \\ 0 & otherwise. \end{cases} \tag{2}$$

The size of the ranked list can assume different values depending on the desired analysis. In the proposed method, the matrix W is defined assuming $m \leq k$, for a local neighborhood analysis, and $m = L$ for a more comprehensive collection sub-set. Notice that, since the matrix W has dimension of $n \times n$, both k and L values are much smaller than n, i.e, $k, L \ll n$. Therefore, the matrix W is very sparse. This property is exploited for defining an efficient algorithm, which computes operations that are equivalent to a matrix multiplication considering W.

4.2 Reciprocal Rank Normalization

While most of similarity pairwise measures are symmetric, the same does not occur for rank measures. In other words, an image img_i well ranked for a query img_j does not imply that img_j is well ranked for query img_i. However, the benefits of improving the symmetry of the k-neighborhood relationship are remarkable in image retrieval applications [12]. Thus, a simple rank normalization procedure is conducted before the rank diffusion process. The reciprocal references among all ranked lists at top-L positions are considered, such $m = L$. For this, the similarity matrix W_L is used and its asymmetry is exploited. The value of w_{ij} is defined considering the position of img_j in the ranked list τ_i, while w_{ji} considers the position of img_i in τ_j. Therefore, a normalized rank similarity matrix $\bar{R_L}$ can be defined as the sum of the original matrix W with its transposed:

$$\bar{R} = W_L + W_L^T. \tag{3}$$

Based on the matrix $\bar{R}$, a rank normalized distance $\bar{\rho}$ is defined as:

$$\bar{\rho}(i,j) = \frac{1}{1+\bar{r}_{ij}}, \tag{4}$$

where $\bar{r}_{ij} \in \bar{R}$.

In the following, all the ranked lists are updated according to the normalized distance, using a stable sorting algorithm. In this way, all similarity scores defined as 0 in the matrix $\bar{R}$ have their distance changed to 1. In these cases, after the execution of the stable sorting algorithm, the previous order of ranked lists are maintained.

This update gives rise to a new set of ranked lists $\bar{\mathcal{T}}$, used as input for the next steps of the proposed algorithm.

4.3 Iterative Rank Diffusion

The proposed rank diffusion approach is defined by an iterative update of similarity information encoded into a matrix P. The update at each iteration is computed according to a rank similarity matrix W of increasing neighborhood size. The central idea consists in spreading the similarity information through P considering initially a small neighborhood, which is gradually expanded over iterations. Therefore, the number of iterations is defined proportionally to the neighborhood size.

Formally, let $^{(t)}$ denote the current iteration and let s be a constant value, which defines the initial neighborhood size. The rank similarity matrix at a given iteration t is defined as:

$$W^{(t)} = W_{s+t}. \tag{5}$$

where the size of ranked lists $m = (s + t)$. The value of $s = 2$ can be used as a default starting value, or s can be manually defined. The initialization of matrix P is defined considering $t = 0$, and therefore:

$$P^{(0)} = W_s. \tag{6}$$

Given the asymmetry rank relationships, a normalization similarity value is computed proportionally to the accumulated rank similarity of each image. The normalization is defined for matrices W and P, respectively as

$$\bar{W_{ij}}^{(t)} = \frac{W_{ij}^{(t)}}{\sum_{c=1}^{n} W_{jc}^{(t)}}, \tag{7}$$

and

$$\bar{P}_{ij}^{(t)} = \frac{P_{ij}^{(t)}}{\sum_{c=1}^{n} P_{jc}^{(t)}}. \tag{8}$$

The iterative diffusion step is defined in terms of the multiplication of the normalized matrices $\bar{P}$ and $\bar{W}$. At each iteration, a larger neighborhood is considered in $\bar{W}$ and disseminated along P:

$$P^{(t+i)} = \bar{P}^{(t)} \bar{W^{(t)}}^{T}, \tag{9}$$

where i indicates the increment (its default value is 1). The process is iteratively executed while $t \leq (k - s)$, where k is a parameter that defines the size of the neighborhood considered.

4.4 Reciprocal Rank Diffusion

The diffusion step defined in Equation 9 considers the transposed matrix $\bar{W^{(t)}}^{T}$. In this way, the similarity of a given image img_i to other images is updated according to the ranked list τ_i and is encoded in the similarity matrix.

However, the reciprocal similarity information should also be considered. With this purpose, after the iterative rank diffusion, a self-diffusion step is defined as:

$$P_r = \bar{P}^{(k-s)} \bar{P}^{(k-s)}, \tag{10}$$

where $\bar{P}^{(k-s)}$ represents the last matrix computed by the iterative diffusion after normalization.

4.5 Rank Diffused Distance

Finally, a new rank diffused distance ρ_d is computed inversely proportional to the reciprocal similarity matrix P_r:

$$\rho_r(i,j) = \frac{1}{1+P_{r_{ij}}} \tag{11}$$

Based on the distance ρ_r, the new and more effective set of ranked lists $\mathcal{T}_r$ is computed using a stable sorting algorithm. As for the rank normalization step, images, which present similarity values equal to 0, maintain the previous order in the ranked lists.

5. RANK AGGREGATION

Different features often encode distinct and complementary visual information extracted from images. Therefore, if a feature produces effective rankings by itself and is complementary (heterogeneous) to other features, then it is expected that a higher search accuracy can be achieved by combining them [42].

In this work, a rank aggregation approach is presented for combining different rankings using the proposed rank diffusion process. The rank diffusion is performed in two stages: first, for each descriptor in isolation and in the next, considering a fused set of ranked lists.

Let $\mathcal{D} = \{D_1, D_2, \ldots, D_d\}$ be a set of different image descriptors and let $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_d\}$ be their respective set of ranked lists. The rank diffusion process is computed for each set $\mathcal{T}_i$, in order to compute a matrix P_r (Equation 10). In the following, a fused matrix P_f is defined as:

$$P_f = \sum_{j \in \mathcal{D}} P_{r_j}. \tag{12}$$

Based on P_f, a new distance ρ_f is computed:

$$\rho_f(i,j) = \frac{1}{1+P_{f_{ij}}}. \tag{13}$$

A fused set of ranked lists $\mathcal{T}_f$ is computed using the distance ρ_f. Finally, we aim at exploiting the contextual information of the fused set of ranked lists $\mathcal{T}_f$. Once the set $\mathcal{T}_f$ presents the same structure of a set obtained for a single descriptor, it is submitted to the regular rank diffusion process, giving rising to a final set $\mathcal{T}_r$.

6. EXPERIMENTAL EVALUATION

The proposed method was evaluated on shape and natural image retrieval tasks, considering the MPEG-7 [17] and UKBench [22] datasets, which are commonly used as benchmark for image retrieval and post-processing methods. All images of each dataset are used as query images. Regarding parameters, we used s=5, i=5, and k=20 for the MPEG-7 [17] dataset and s=2, i=2, and k=6 for the UKBench [22] dataset, due to the small number of images per class.

6.1 Shape Retrieval

The shape retrieval experiments were condcuted on the MPEG-7 [17], which is a well-known shape dataset, composed of 1400 shapes divided into 70 classes. Six shape descriptors were considered and the *Mean Average Precision* (MAP) was used as effectiveness measure.

Table 1 presents the experimental results for shape retrieval experiments. Statistical paired t-tests were conducted comparing the results before and after the use of the proposed algorithm. Different values of L are considered: $L = 400$ and the whole ranked list. Positive gains with statistical significance can be observed for all descriptors, reaching high effectiveness gains up to +40.72%. The effectiveness results obtained for partial and the entire ranked lists are very similar, demonstrating that only a sub-set of ranked lists is enough to obtain high effectiveness gains.

Results for rank aggregation tasks are also presented considering the combination of the three best descriptors. The relative gains are computed based on the descriptor with the lowest MAP score. As it can be observed, high effective results are also obtained, reaching 99.78% for CFD+AIR.

Table 1: Rank Diffusion for shape retrieval tasks (MAP as score).

Descriptor	**Orig. MAP**	$L = 400$			*Full L*		
		Rank Diff.	**Gain**	**Stat. Sig.**	**Rank Diff.**	**Gain**	**Stat. Sig.**
SS [8]	37.67%	52.17%	+38.49%	•	53.01%	+40.72%	•
BAS [1]	71.52%	82.36%	+15.16%	•	83.03%	+16.09%	•
IDSC [18]	81.70%	90.89%	+11.25%	•	91.09%	+11.49%	•
CFD [25]	80.71%	93.75%	+16.16%	•	94.17%	+16.68%	•
ASC [19]	85.28%	92.98%	+9.03%	•	93.07%	+9.13%	•
AIR [10]	89.39%	97.98%	+9.61%	•	97.97%	+9.60%	•
CFD+ASC	-	99.29%	+23.02%	•	99.25%	+22.97%	•
ASC+AIR	-	99.57%	+16.76%	•	99.58%	+16.77%	•
CFD+AIR	-	99.78%	+23.63%	•	99.79%	+23.64%	•

6.2 Natural Image Retrieval

We evaluate the proposed method in natural image retrieval tasks considering the University of Kentucky Recognition Benchmark – UKBench [22]. The UKBench dataset has a total of 10,200 images, composed of 2,550 objects or scenes, where each object/scene is captured 4 times from different viewpoints. The small number of images per class constitutes a challenging scenario for unsupervised learning approaches. Various distinct features are used, including color and color/texture features available on LIRE framework [20]. Local features are considered based on Vocabulary Tree (VOC) [36]. Convolution neural networks (CNN) features based on Caffe [13] framework are also considered.

Table 2 presents the effectiveness results considering the UKBench [22] dataset. The N-S score is used as effectiveness measure, varying between 1 and 4. This score corresponds to the number of relevant images among the first four image returned (the highest achievable score is 4). We can observe significant improvements for N-S scores. Notice, for example, the Caffe [13] convolutional neural network, which is improved from 3.31 to 3.61. The results are even more impressive considering the rank aggregation tasks.

6.3 Comparison with Other Approaches

The proposed method is also evaluated in comparison with various other state-of-the-art approaches and recently proposed retrieval approaches. Table 3 presents the results of Rank Diffusion method on the UKBench [22] dataset in comparison with recent retrieval approaches. Table 4 presents the obtained results on the MPEG-7 [17] dataset in comparison with various other state-of-the-art post-processing methods. The bull's eye score, which counts the matching shapes within the top-40 ranked images, is used as evaluation measure. As it can be observed, the effectiveness results of the proposed method compares favorably with the recent retrieval approaches.

Table 2: Rank Diffusion on the UKBench [22] dataset.

Descriptor	**Original N-S**	**Rank Diffusion**
CEED-SPy [4, 20]	2.81	3.10
FCTH-SPy [5, 20]	2.91	3.19
SCD [21]	3.15	3.35
ACC-SPy [11, 20]	3.25	3.51
CNN-Caffe [13]	3.31	3.61
ACC [20]	3.36	3.60
VOC [36]	3.54	3.72
VOC+ACC	-	3.90
VOC+CNN	-	3.90
ACC+CNN	-	3.87
VOC+ACC+CNN	-	**3.94**

Table 3: Retrieval comparison on the UKBench [22] dataset.

N-S scores for recent retrieval methods					
Zheng *et al.* [43]	Qin *et al.* [28]	Wang *et al.* [34]	Zhang *et al.* [41]	Zheng *et al.* [42]	Xie *et al.* [37]
3.57	3.67	3.68	3.83	3.84	3.89

N-S scores for the Rank Diffusion method			
Rank Diff.: ACC+CNN	Rank Diff. VOC+CNN	Rank Diff. VOC+ACC	Rank Diff. VOC+ACC+CNN
3.88	3.90	3.90	**3.94**

Table 4: Comparison on the MPEG-7 [17] dataset.

Shape Descriptors		
CFD [25]	-	84.43%
IDSC [18]	-	85.40%
ASC [19]	-	88.39%
AIR [10]	-	93.67%
Post-Processing Methods		
Algorithm	**Descriptor(s)**	**Score**
Graph Transduction [38]	IDSC	91.00%
Locally C. Diffusion Process [39]	IDSC	93.32%
Shortest Path Propagation [35]	IDSC	93.35%
Locally C. Diffusion Process [39]	ASC	95.96%
Rank Diffusion	**CFD**	**96.19%**
Tensor Product Graph [40]	AIR	99.99%
Generic Diffusion Process [9]	AIR	100%
Neighbor Set Similarity [2]	AIR	100%
Rank Diffusion	**AIR**	**100%**

7. CONCLUSIONS

Post-processing procedures based on unsupervised learning approaches have been established as indispensable tools for improving the effectiveness of CBIR systems. Since diffusion process methods require high computational efforts, rank-based approaches attracted a lot of research attention recently to circumvent their limitations. In this paper, a novel rank diffusion method is proposed exploiting characteristics of both diffusion and rank-based approaches. A rigorous experimental evaluation demonstrated the effectiveness of the proposed approach. Future work focuses on the deep investigation of contextual information encoded in the rank similarity matrix.

8. ACKNOWLEDGMENTS

The authors are grateful to FAPESP (grants 2013/08645-0 and 2013/50169-1), CNPq (grants 306580/2012-8 and 484254/2012-0), CAPES, AMD, and Microsoft Research.

9. REFERENCES

[1] N. Arica and F. T. Y. Vural. BAS: a perceptual shape descriptor based on the beam angle statistics. *Pattern Recognition Letters*, 24(9-10):1627–1639, 2003.

[2] X. Bai, S. Bai, and X. Wang. Beyond diffusion process: Neighbor set similarity for fast re-ranking. *Information Sciences*, 325:342 – 354, 2015.

[3] P. Brodatz. *Textures: A Photographic Album for Artists and Designers*. Dover, 1966.

[4] S. A. Chatzichristofis and Y. S. Boutalis. Cedd: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval. In *ICVS*, pages 312–322, 2008.

[5] S. A. Chatzichristofis and Y. S. Boutalis. Fcth: Fuzzy color and texture histogram - a low level feature for accurate image retrieval. In *WIAMIS*, pages 191–196, 2008.

[6] Y. Chen, X. Li, A. Dick, and R. Hill. Ranking consistency for image matching and object retrieval. *Pattern Recognition*, 47(3):1349 – 1360, 2014.

[7] R. R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5 – 30, 2006.

[8] R. da S. Torres and A. X. Falcão. Contour Salience Descriptors for Effective Image Retrieval and Analysis. *Image and Vision Computing*, 25(1):3–13, 2007.

[9] M. Donoser and H. Bischof. Diffusion Processes for Retrieval Revisited. In *CVPR*, pages 1320–1327, 2013.

[10] R. Gopalan, P. Turaga, and R. Chellappa. Articulation-invariant representation of non-planar shapes. In *ECCV*, volume 3, pages 286–299, 2010.

[11] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih. Image indexing using color correlograms. In *CVPR*, pages 762–768, 1997.

[12] H. Jegou, C. Schmid, H. Harzallah, and J. Verbeek. Accurate image search using the contextual dissimilarity measure. *IEEE TPAMI*, 32(1):2–11, 2010.

[13] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.

[14] J. Jiang, B. Wang, and Z. Tu. Unsupervised metric learning by self-smoothing operator. In *ICCV*, pages 794–801, 2011.

[15] P. Kontschieder, M. Donoser, and H. Bischof. Beyond pairwise shape similarity analysis. In *ACCV*, pages 655–666, 2009.

[16] V. Kovalev and S. Volmer. Color co-occurence descriptors for querying-by-example. In *ICMM*, page 32, 1998.

[17] L. J. Latecki, R. Lakmper, and U. Eckhardt. Shape descriptors for non-rigid shapes with a single closed contour. In *CVPR*, pages 424–429, 2000.

[18] H. Ling and D. W. Jacobs. Shape classification using the inner-distance. *IEEE TPAMI*, 29(2):286–299, 2007.

[19] H. Ling, X. Yang, and L. J. Latecki. Balancing deformability and discriminability for shape matching. In *ECCV*, volume 3, pages 411–424, 2010.

[20] M. Lux. Content based image retrieval with LIRe. In *ACM MM '11*, 2011.

[21] B. Manjunath, J.-R. Ohm, V. Vasudevan, and A. Yamada. Color and texture descriptors. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):703–715, 2001.

[22] D. Nistér and H. Stewénius. Scalable recognition with a vocabulary tree. In *CVPR*, volume 2, pages 2161–2168, 2006.

[23] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE TPAMI*, 24(7):971–987, 2002.

[24] D. C. G. Pedronette, J. Almeida, and R. da S. Torres. A scalable re-ranking method for content-based image retrieval. *Information Sciences*, 265(1):91–104, 2014.

[25] D. C. G. Pedronette and R. da S. Torres. Shape retrieval using contour features and distance optmization. In *VISAPP*, volume 1, pages 197 – 202, 2010.

[26] D. C. G. Pedronette and R. da S. Torres. Image re-ranking and rank aggregation based on similarity of ranked lists. *Pattern Recognition*, 46(8):2350–2360, 2013.

[27] D. C. G. Pedronette and R. da S. Torres. Unsupervised manifold learning using reciprocal knn graphs in image re-ranking and rank aggregation tasks. *Image and Vision Computing*, 32(2):120–130, 2014.

[28] D. Qin, S. Gammeter, L. Bossard, T. Quack, and L. van Gool. Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors. In *CVPR*, pages 777 –784, june 2011.

[29] X. Shen, Z. Lin, J. Brandt, S. Avidan, and Y. Wu. Object retrieval and localization with spatially-constrained similarity measure and k-nn re-ranking. In *CVPR*, pages 3013 –3020, june 2012.

[30] R. O. Stehling, M. A. Nascimento, and A. X. Falcão. A compact and efficient image retrieval approach based on border/interior pixel classification. In *CIKM*, pages 102–109, 2002.

[31] M. J. Swain and D. H. Ballard. Color indexing. *International Journal on Computer Vision*, 7(1):11–32, 1991.

[32] B. Tao and B. W. Dickinson. Texture recognition and image retrieval using gradient indexing. *JVCIR*, 11(3):327–342, 2000.

[33] J. van de Weijer and C. Schmid. Coloring local feature extraction. In *ECCV*, pages 334–348, 2006.

[34] B. Wang, J. Jiang, W. Wang, Z.-H. Zhou, and Z. Tu. Unsupervised metric fusion by cross diffusion. In *CVPR*, pages 3013 –3020, 2012.

[35] J. Wang, Y. Li, X. Bai, Y. Zhang, C. Wang, and N. Tang. Learning context-sensitive similarity by shortest path propagation. *Pattern Recognition*, 44(10-11):2367–2374, 2011.

[36] X. Wang, M. Yang, T. Cour, S. Zhu, K. Yu, and T. Han. Contextual weighting for vocabulary tree based image retrieval. In *ICCV'2011*, pages 209–216, Nov 2011.

[37] L. Xie, R. Hong, B. Zhang, and Q. Tian. Image classification and retrieval are one. In *ACM ICMR '15*, pages 3–10, 2015.

[38] X. Yang, X. Bai, L. J. Latecki, and Z. Tu. Improving shape retrieval by learning graph transduction. In *ECCV*, volume 4, pages 788–801, 2008.

[39] X. Yang, S. Koknar-Tezel, and L. J. Latecki. Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval. In *CVPR*, pages 357–364, 2009.

[40] X. Yang, L. Prasad, and L. Latecki. Affinity learning with diffusion on tensor product graph. *IEEE TPAMI*, 35(1):28–38, 2013.

[41] S. Zhang, M. Yang, T. Cour, K. Yu, and D. Metaxas. Query specific rank fusion for image retrieval. *IEEE TPAMI*, 37(4):803–815, April 2015.

[42] L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, and Q. Tian. Query-adaptive late fusion for image search and person re-identification. In *CVPR'2015*, 2015.

[43] L. Zheng, S. Wang, and Q. Tian. Lp-norm IDF for scalable image retrieval. *IEEE TIP*, 23(8):3604–3617, Aug 2014.