2015-12

Combining diversity queries and visual mining to improve content-based image retrieval systems: the DiVI method

IEEE International Symposium on Multimedia, 2015, Miami.
http://www.producao.usp.br/handle/BDPI/49671

Combining Diversity Queries and Visual Mining to Improve Content-Based Image Retrieval Systems: The DiVI Method

Lucio F. D. Santos*, Rafael L. Dias†, Marcela X. Ribeiro†,
Agma J. M. Traina* and Caetano Traina Jr*
*Department of Computer Science, University of São Paulo, Brazil
†Department of Computing, Federal University of São Carlos, Brazil
*{luciodb, agma, caetano}@icmc.usp.br
†{rafael.dias,marcela}@dc.ufscar.br

***Abstract*—This paper proposes a new approach to improve similarity queries with diversity, the Diversity and Visually-Interactive method (DiVI), which employs Visual Data Mining techniques in Content-Based Image Retrieval (CBIR) systems. DiVI empowers the user to understand how the measures of similarity and diversity affect their queries, as well as increases the relevance of CBIR results according to the user judgment. An overview of the image distribution in the database is shown to the user through multidimensional projection. The user interacts with the visual representation changing the projected space or the query parameters, according to his/her needs and previous knowledge. DiVI takes advantage of the users' activity to transparently reduce the semantic gap faced by CBIR systems. Empirical evaluation show that DiVI increases the precision for querying by content and also increases the applicability and acceptance of similarity with diversity in CBIR systems.**

***Keywords*-Content-Based Image Retrieval; Semantic Gap; Similarity With Diversity; Visual Data Mining;**

I. Introduction

As the volume of images produced and gathered exponentially grows in time, the demand for tools to retrieve images by similarity becomes increasingly popular [1], [2].

Content-Based Image Retrieval (CBIR) systems retrieve images that are similar to a query image using its visual content, taking into account the distance (similarity) from the visual features of the query image to the other images in the dataset. Images can be associated to several visual features such as color, shape and texture. Typically, CBIR systems represent each image as a feature vector, which is a numerical representation summarizing the visual features, known as image description, automatically obtained by image processing algorithms. However, images perceived as similar by humans often do not have similar feature vectors, and vice-versa. The difference between the numeric representation found by image processing algorithms and the human cognition is called the *semantic gap*.

Previous studies have shown that including the users into the retrieval process is a way of reducing the semantic gap [2]. An approach commonly used to deal with the semantic gap is Relevance Feedback (RF). Generally, it requires that the users feed the system with their grades about the query results and asks the system to re-issue the query. However, RF requires both a lot of computational resources and the users' ability to properly tune the system, which may be cumbersome or even not possible at all.

Another way to reduce the semantic gap is speeding up the user interaction, using the computational power to more objectively evaluate the relationship among images in the entire dataset. This is important as both the description of the images and the process to retrieve the images most similar to the query performed by conventional CBIR systems are mostly unclear to the user, which usually see those processes as a black box. Visual Data Mining (VDM) techniques may help by tilting the distance evaluation to produce results closer to the human cognition, as it makes possible to take advantage of the user interaction in the retrieval process. However, to the best of our knowledge, just few works exist in literature [3], [4], [5] regarding applying VDM to CBIR-based tools, and none of them aim at helping the user to interact with those systems.

Another way to approximate the results to the user's needs is adding a diversity factor in the query process, besides using just similarity. The purpose aims at obtaining elements in the result that are not only similar to the query element, but also diverse among themselves. A more heterogeneous result would provide more knowledge than a too uniform one, since images from different categories may be retrieved, giving to the user a better perspective of the possible results from the query. Recent studies [6], [1], [7] employ diversity to reduce the semantic gap in image retrieval. However, the intuition of how diversity is applied in image datasets has not been explored yet, precluding the user from getting the advantage of receiving a diversified answer.

This paper presents the Diversity and Visually-Interactive method (DiVI), a novel approach to integrate diversity and visualization in CBIR systems. It provides a multidimensional projection to visualize the image space formed by the extracted image features from the entire dataset. It allows interacting with the visualizations by navigating and bending the data space towards his/her expectations, making the

978-1-5090-0379-2/15 $31.00 © 2015 IEEE
DOI 10.1109/ISM.2015.115

user's participation more active in the process and steering the results according to his/her needs.

The remainder of this paper is structured as follows. Section II summarizes the main concepts and related works. Section III presents our technique, while Section IV details the experimental evaluation performed over a public image dataset, which describes and analyzes the results achieved. Finally, Section V presents the conclusions and future works.

II. Background and Related Works

There are two types of similarity queries performed by CBIR systems: the similarity range and the k-nearest neighbor ones. The similarity range query (R_q) finds all images within a given query distance *r* from the query image, while the *k*-Nearest Neighbor query (k-NN$_q$) retrieves the *k* images most similar to the query one. However, querying a massive image database employing either k-NN$_q$ or R_q predicates may often retrieve images too similar among themselves, increasing the analysis effort required from the specialist.

A solution is to take into account a diversity degree in the result set of the similarity predicates. The most common method to provide diversity is solving a bi-criteria optimization problem, where similarity and diversity compete with each other, ruled by a *trade-off* parameter (λ), known as the diversity preference [1]. Diversity is induced on the traditional similarity results by re-ranking the elements based on the bi-criteria objective function, thus those methods result in an NP-hard problem [7].

An interesting way to overcome the semantic gap is using Visual Data Mining (VDM) techniques, which includes a knowledge extraction step that unites computational efficiency with human judgment. Different visualization and interaction strategies are being studied to facilitate the analysis and exploration of image datasets and their information by the users [8], [3], [4]. The VDM techniques are classified according to the information to be displayed, the visualization technique, and the interaction and distortion techniques employed [9]. In CBIR systems, the mostly used visualization techniques are based on hierarchical and multidimensional projection, and the techniques of interaction and distortion depend on the user's needs. An important factor concerning the viability of these tools is to facilitate the data exploration to make a better use of the human visual ability.

Typically, multidimensional projection techniques provide a way to map the data from an *n*-dimensional into a *d*-dimensional space, $d = 1, 2, 3 \ldots$, yet preserving as much as possible the distance relationship among elements in the original space. A multidimensional projection moves the elements in a d-dimensional space and checks how well the distances between them can be reproduced in this space. It minimizes $\Phi = \sum((d_{ij} - f(\delta_{ij}))^2$ where, d_{ij} stands for the projected space distances and δ_{ij} stands for the original distances. The expression $f(\delta_{ij})$ indicates a nonmetric, monotone transformation of the original distances, which reproduces the original distances between the elements.

In this paper, we employ two multidimensional projection techniques in our proposed method: FASTMAP [10] and LAMP [8]. The FASTMAP method finds N points to map N elements into a $z-d$ space, where d is the dimension of the dataset elements, whose Euclidean distances match the distances of a given $N \times N$ distance matrix M. The elements are treated as points into an n-dimensional space, and projected on z mutually orthogonal directions. The LAMP method maps each point x using the affine transformation $f_x(p) = pM + t$ that minimizes:

$$\sum_{i=1} \alpha_i (f_x(x_i) - y_i)^2 \tag{1}$$

$$\alpha_i = \frac{1}{(x_i - x)^2} \tag{2}$$

where x_i is a control point, y_i is its mapping in the projected space and x is an object instance to be mapped. The LAMP goal is to allow the user to define control points positions that guide the final projection and perform local projections guided by the control points. This representation takes advantage of the human visual ability to identify relations or patterns in datasets.

Few works in the literature applied visualization to CBIR. In [3], they employ the Pathfind network technique to visualize the relationship among TV images. In [4], an image visualization system to analyze the feature vector quality in a CBIR system was proposed. These techniques employ a limited, static visualization of the image dataset. The Projection Explorer system for Images (PEx-Image) [5] proposes the visualization of the features extracted from image datasets through a multidimensional projection and hierarchical visualization. Its main idea is to combine multidimensional projections and hierarchical visualizations, automatic feature selection and user interaction to explore image datasets distribution and classification. However, this approach only allows comparison by similarity, without considering diversity.

III. The DiVI Method

Aimed at reducing the semantic gap in CBIR systems, our proposed Diversity and Visually-Interactive (DiVI) method combines diversity and visual data mining techniques to improve retrieval efficiency. It includes the user into the processing path, to interactively distort the search space in the image description process, forcing the elements that he/she considers more similar to be closer and elements considered less similar to be farther in the search space. Thus, DiVI allows inducing in the space the intuitive perception of similarity lacking in the numeric evaluation of the distance function. It also allows the user to express his/her diversity preference for a query, reducing the effort to analyze the result when too much similar images are returned.

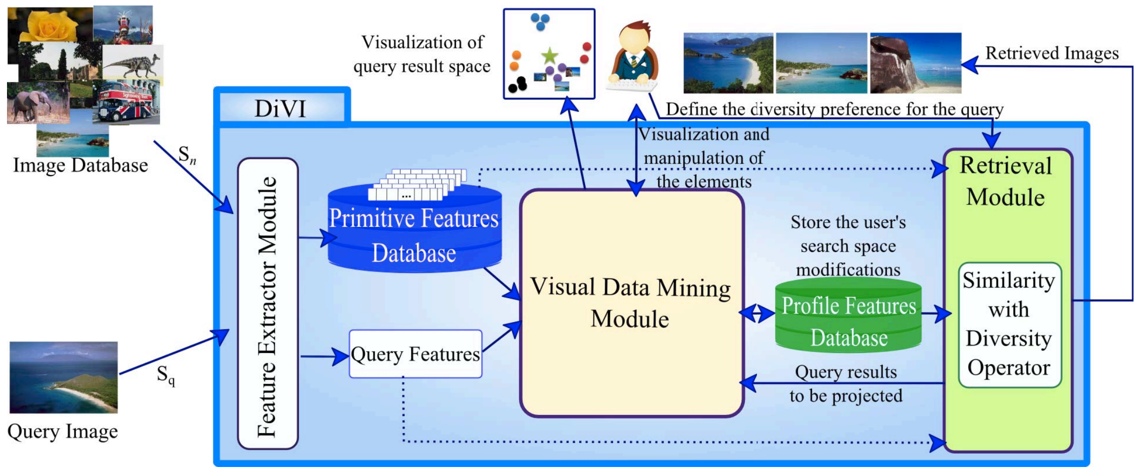


Figure 1: Pipeline of DiVI processing embedded in a CBIR-based tool.

A. *Detailed Method*

An overview of DiVI either using VDM techniques (solid arrows) or not (dotted arrows) is shown in Figure 1. Upon starting a DiVI-enabled CBIR, the user provides the usual parameters, such as the query image and the desired number k of images to be retrieved. Thereafter, both the image database and the query image are compared based on the visual features as color, texture or shape extracted by the Feature Extractor Module (FEM). The extracted feature vectors are stored in the Primitive Feature (PF) database. Notice that the first phase of image description is processed in the same way as in traditional CBIR.

When the user does not use VDM techniques to interact with the system, the features stored as PF are used by the data retrieval module, which includes diversity-enabled similarity operators. It also allows the user to define his/her own *diversity preference* (λ) for the query, as shown in Figure 1 (following the flow with dotted arrows). When the VDM module is activated, the initial projection of the search space is produced, and the user can manipulate and insert his/her knowledge about similarity (the following solid arrows in Figure 1). In order to reduce the complexity and the effort spent by the user, samples that represent each category of the image dataset (called as Control Points - CP) along with the projection of the query image are pre-selected using the FASTMAP technique. The user manipulates the projections of the CP and of the query image by adjusting and positioning them in the projected space according to his/her own judgment. The user's modification on the search space are stored in a Profile Features database persisting this information for other sessions.

The Profile Features database provides the feature vector to the Retrieval Module, which re-issues the query, now over the query tilted space. A new projection is generated using as reference the distortion of the original space stored in the Profile Features database with the query results retrieved by the Retrieval Module using a multidimensional projection technique based on LAMP. Comparing the original and the tilted spaces helps improving the user's understanding about the query. Following, we detail how diversity and VDM techniques are applied to the query process.

1) Similarity with diversity: Diversity is induced as a bi-criteria optimization problem. Formally, let $S = \{s_1, ..., s_n\}$ be the set of n stored elements; q a query image and S' be the candidate subset of m elements obtained by the R_q with a similarity threshold r to the query image; λ ($0 \leq \lambda \leq 1$) be the *trade-off* between similarity and diversity, called the *diversity preference*; and k be the number of desired returning elements ($k \leq n$). Thus, the diversity for a query centered at q with result set R is given by:

$$R = argmax_{S' \subseteq S, k=|S'|} \mathcal{F}(q, S') \ , \tag{3}$$

where the score function $\mathcal{F}(q, S')$ for each candidate subset S' is defined as:

$$\mathcal{F}(q, S') = (k-1)(1-\lambda) \cdot sim(q, S') + 2\lambda \cdot div(S') \ . \tag{4}$$

Notice that, there are two special cases for the diversity preference parameter λ in (4). When $\lambda = 0$, the final result depends only on the query image q, reducing the problem to find the most similar images to q. The second case is when $\lambda = 1$ and the similarity to q does not play any role in the final result, reducing the problem to find the most diverse images of a subset S'. Thus, the user can control the λ parameter sliding from only similarity answers to similarity with diversity to just diversity.

The similarity represented by Equation (5) verifies the distance between the query image and the other images in the subset.

$$sim(q, S') = \sum_{i=1}^{k} \delta_{sim}(q, s_i), s_i \in S' \ , \tag{5}$$

(a) Initial projection without user manipulation.

(b) Initial projection after user manipulation with instance thumbnail visualization.

Figure 2: Examples of handling instances in the initial projection in the VDM module.

The $\delta_{sim}(q, s_i)$ function is monotonically increasing, i.e., higher values implies that the image $s_i \in S'$ is more relevant to the query q. The diversity is verified through Equation (6), which calculates the distance between all images in the subset using the distance function $\delta_{div}(s_i, s_j)$.

$$div(S') = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \delta_{div}(s_i, s_j), s_i, s_j \in S' \ . \qquad (6)$$

2) Visual Data Mining Techniques: We use Visual Data Mining techniques to visualize and interact with instances at the user's discretion. The multidimensional projection performed by FASTMAP is used as the initial projection with some samples of the image dataset (control points), along with the query image. At this stage, the user can manipulate the placement of such instances and steer the search space. Figure 2 shows an example of handling the instances provided by the VDM module, where the star represents the query image and each point is a representation of an image of the dataset. Figure 2 (a) shows the initial projection without user manipulation. The star stands for the query image projection and the points represent the images in the dataset. The user manipulates the projection by moving either the points or the star point. Figure 2(b) shows the projection after the user manipulation. The colors indicate the class of each image, retrieved from its metadata. The user may ask for projected image thumbnails to assist the projected space manipulation.

After the space manipulation, a final projection is calculated. The information regarding position of each projected instance is then used to retrieve the answer of similarity with diversity queries. The results guided by the user, the VDM techniques and the results obtained in the orignal n-dimensional space are employed to generate further visualizations. To evince the advantages of the user's guidance in the query process, additional features can be shown to highlight the factors involved in this process, as illustrated in Figure 3. In the tilted projection, it is possible to ask for: the visualization of the centroids of each class/category (Figure 3(a)); the control points used to calculate the final projection (Figure 3(b)); the results using the projection space (Figure 3(c)) and the n-dimensional space (Figure 3(d)).

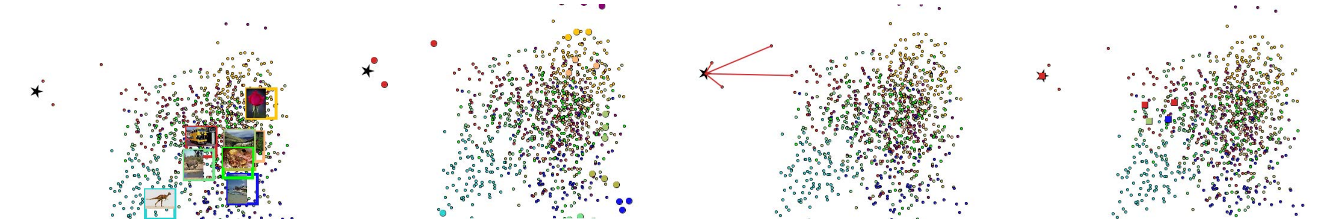
DiVI fosters reducing the semantic gap through a more active participation of the user in the query process, making it possible to check which images were considered in the query, as well as the proximity (similarity) in which they are positioned in the search space. Moreover, the user can modify the diversity preference and see how the images change in the projected space for each query result, improving her/his understanding of the data space and of the diversity factor in the similarity query.

IV. Experiments

Aimed at evaluating the ability of the DiVI method to execute similarity queries, we implemented our DiVI proposal in a CBIR prototype coded in Java EE. We used a variety of image datasets publicly available to perform the experimental evaluation. However, due to space limitations, here we present only the results obtained from one representative and well-known dataset, the `Corel` image dataset [11]. It is composed of 1,000 images of different scenes and objects classified into 10 different categories: Africa; Beach; Building; Bus; Dinosaur; Elephant; Flower; Horse; Mountain and Food. The image feature vectors were obtained using the MPEG-7 Color Layout extractor defined in the ISO/IEC 15938 standard (multimedia content description interface). The comparison between the feature vectors were made using the Euclidean distance function.

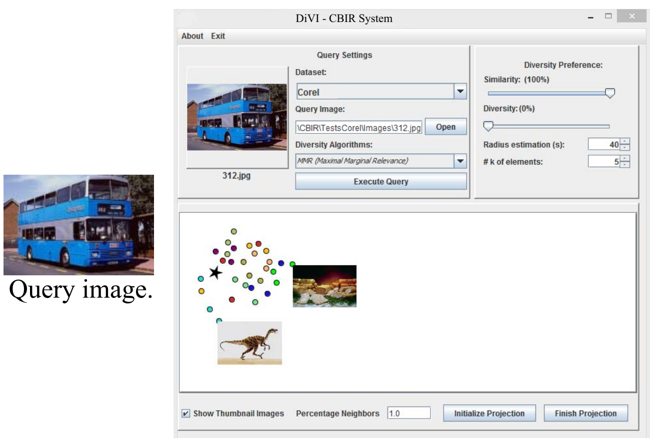
The idea is to evaluate the comprehensiveness of the results found by our proposed method as well as the benefits of adding VDM techniques to a retrieval system. We presented our approach to seven specialists in CBIR to evaluate our prototype. They performed similarity queries with and without diversity using both the primitive features automatically extracted and using the proposed VDM module to adjust the similarity perception. The specialists stated that DiVI is indeed capable of presenting more information about the relationships among the images, yet being able to improve the average precision even when the results are subjective, making the image description more personalized to each user.

To assess the contribution, we request the users to look for the 5 images most similar to the query one, taking into account the similarity of the provided image over the `Corel` dataset, using DiVI. An example is shown in Figure 4, the query image was initially classified as in the Bus category (label 4). The result obtained using just the extracted features and a similarity algorithm is shown in Figure 4 (b). As it can be seen, only two images belong to the same category of the query image (label 4 between parenthesis), and the three remaining images belong to different categories (label 5, 9



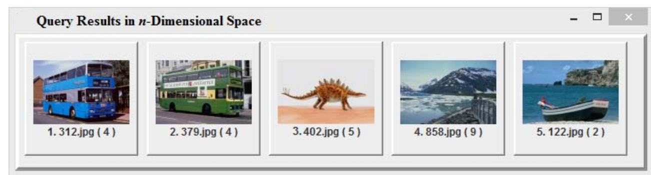
(a) Centroids visualization of each class/category. (b) Control points represented by highlighted circles. (c) Results in the projected space, highlighted by lines connecting to the query image (star). (d) Results in the n-dimensional space, represented by squares.

Figure 3: Visualization and interaction techniques employed in the final projection.



Query image.

(a) Initial Projection.



(b) Query result with solely similarity in n-space.

Figure 4: Results (just similarity) for `Corel` dataset in n-dimensional space (extracted features) activating the VDM module.

and 2) resulting in 40% of precision. Figure 4 (a) presents the Graphical User Interface (GUI) in charge of querying and presenting the initial projection of the search space to the user, showing thumbnails of instance elements. As it can be seen, the image classes are mixed in the search space.

As the query results for the extracted features considering only similarity present different categories to the query image, it can be concluded that the visual features were inadequate to represent images of this class. Thus, the user tilts the search space in accordance to his/her judgment using the VDM module (see Figure 5 (1)). Notice that the image classes now are better grouped. The query results in both projected and n-dimensional space as well as the final projection are shown in Figure 5 (2), (3) and (4), respectively. Through the analysis of the results, one can see that every obtained image is consistent with the category of the query image (precision of 100 %). The visualization of the search space shows the differences among the search in the original n-dimensional space (highlighted by a square in the projected instances) and that one tilted by the user (highlighted by the straight line from the projected instances to the query image in Figure 5 (4)). After the user interactions, the search space modifications are stored in the profile features database and can be used to perform new queries.

Besides the result for the query image presented in window (2) of Figure 5 having 100 % of precision, usually the user is interested in a exploratory query without an homogeneous image result, requiring some diversity. Figure 6 shows the result images considering 40 % of diversity, which now shows different images for the bus category. Notice that, diversity considerations exploits the expressiveness of the retrieval process considering the relationship among the images in the result, instead of only considering similarity to the query image.

V. Conclusions

In this paper, we presented a new method to improve the quality of CBIR systems combining diversity queries and visual data mining techniques to interact with queries, which is called Diversity and Visually-Interactive Method (DiVI). In this method, the results are obtained according to the user needs, which uses visualization and interaction techniques to set the user in a more active role for the query process, enabling a better understanding of the results and

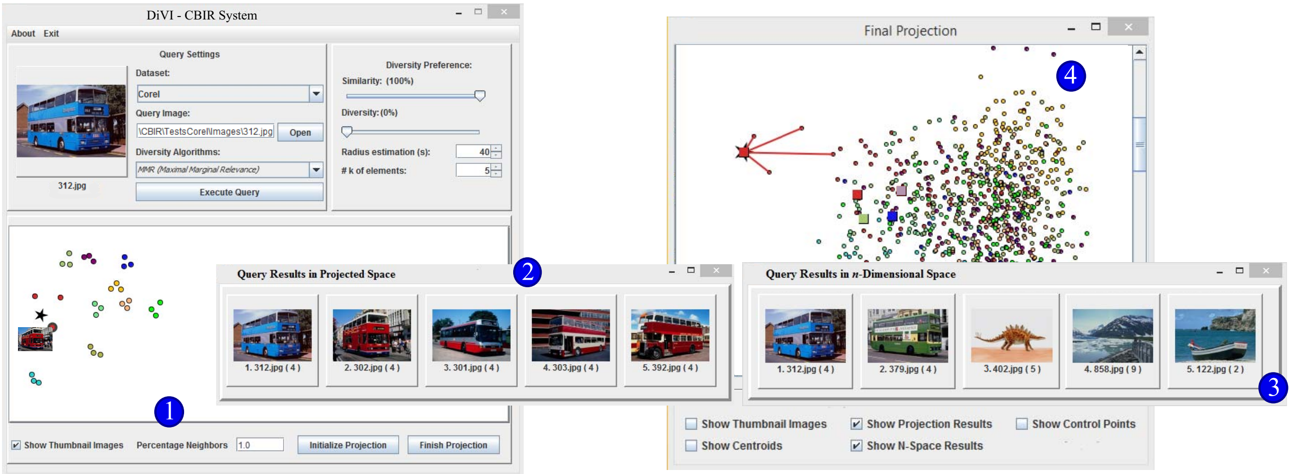


Figure 5: Final projection generated for `Corel` dataset with the query results to both, n-dimensional and user tilted space highlighting the difference between them.

the relationship between images, categories and the extracted features. The empirical evaluation of our methodology indicated that it delivers promising results for image databases. It is able to improve user's understanding and increases the applicability of diversity to reduce the semantic gap in CBIR.

As a future work, we will conduct user-centered analysis regarding the feasibility of using our proposed method in specific domains, such as using medical images to help radiologists in performing diagnosis, as well as in providing more personalized visual features using the tilting space provided by specialists.

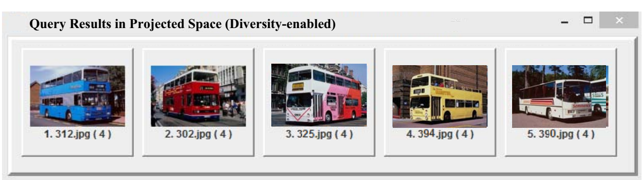


Figure 6: Query result considering 40% of Diversity for `Corel` dataset using the user' search space modifications stored in the profile features database.

Acknowledgment

The authors are grateful to FAPESP, CNPq, CAPES and the Rescuer Project (European Commission Grant 614154 and CNPq/MCTI Grant 490084/2013-3) for their financial support.

References

[1] L. F. D. Santos, W. D. Oliveira, M. R. P. Ferreira, A. J. M. Traina, and C. Traina Jr., "Parameter-free and domain-independent similarity search with diversity," in *SSDBM*, Baltimore, Maryland, 2013, pp. 5:1–5:12.

[2] A. Gallas, W. Barhoumi, and E. Zagrouba, "Negative relevance feedback for improving retrieval in large-scale image collections," in *IEEE ISM 2014*, Dec 2014, pp. 1–8.

[3] G. Gagaudakis and P. Rosin, "Using cbir and pathfinder networks for image database visualisation." in *ICPR*, vol. 1. IEEE Comput. Soc, 2000, pp. 1052–1055.

[4] Q. G. Gang Hu, "An interactive image feature visualization system for supporting cbir study," *Image Analysis and Recognition*, vol. 5627, no. 1, pp. 239–247, 2009.

[5] D. M. Eler, M. Y. Nakazaki, F. V. Paulovich, D. P. Santos, G. F. Andery, M. C. F. Oliveira, J. Batista Neto, and R. Minghim, "Visual analysis of image collections," *The Visual Computer*, vol. 25, no. 10, pp. 923–937, 2009.

[6] R. L. Dias, R. Bueno, and M. X. Ribeiro, "Reducing the complexity of k-nearest diverse neighbor queries in medical image datasets through fractal analysis," in *IEEE CBMS*, 2013, pp. 101–106.

[7] M. Vieira, H. Razente, M. Barioni, M. Hadjieleftheriou, D. Srivastava, C. Traina Jr., and V. Tsotras, "On query result diversification," in *IEEE ICDE 2011*, 2011, pp. 1163–1174.

[8] P. Joia, F. V. Paulovich, D. Coimbra, J. A. Cuminato, and L. G. Nonato, "Local affine multidimensional projection," *IEEE TVCG*, vol. 17, no. 12, pp. 2563–2571, 2011.

[9] D. A. Keim *et al.*, "Information visualization and visual data mining," *IEEE TVCG*, vol. 8, no. 1, pp. 1–8, 2002.

[10] C. Faloutsos and K.-I. Lin, "Fastmap: a fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets," *SIGMOD Rec.*, vol. 24, no. 2, pp. 163–174, May 1995.

[11] J. Z. Wang, J. Li, and G. Wiederhold, "Simplicity: Semantics-sensitive integrated matching for picture libraries," *IEEE TPAMI*, vol. 23, no. 9, pp. 947–963, 2001.