Noname manuscript No.
(will be inserted by the editor)

Class-Specific Metrics for Multidimensional Data Projection Applied to CBIR

Paulo Joia · Erick Gomez-Nieto · João Batista Neto
Wallace Casaca · Glenda Botelho · Afonso Paiva · Luis Gustavo Nonato

Received: date / Accepted: date

Abstract Content-based image retrieval is still a challenging issue due to the inherent complexity of images and choice of the most discriminant descriptors. Recent developments in the field have introduced multidimensional projections to burst accuracy in the retrieval process, but many issues such as introduction of pattern recognition tasks and deeper user intervention to assist the process of choosing the most discriminant features still remain unaddressed. In this paper we present a novel framework to CBIR that combines pattern recognition tasks, class-specific metrics and multidimensional projection to devise an effective and interactive image retrieval system. User interaction plays an essential role in the computation of the final multidimensional projection from which image retrieval will be attained. Results have shown that the proposed approach outperforms existing methods, turning out to be a very attractive alternative for managing image data sets.

Keywords Multidimensional projection · Content-Based Image Retrieval.

1 Introduction

Content-based image retrieval (CBIR) is a basic tool in any computational system aimed at supporting the cataloging and querying of images from large databases such as art collections, photograph archives, and medical diagnosis. Many different mechanisms have been proposed to accomplish the CBIR process, which vary as to the metric used to compare images as well as the way the search results are presented (see [6] for a comprehensive survey). Among the different approaches, multidimensional projection-based CBIR [8] has emerged as one of the most promising method, since it allows for simultaneously querying multiple images while visualizing the results in a two-dimensional manner that enables user interaction.

Despite the advances and good results, existing multidimensional projection (MP) methods have not yet exploited all its potential towards incorporating well-known pattern recognition tools into the CBIR process. For instance, the MP techniques are very flexible in terms of the metric used to measure the similarity between instances of data, thus, the metric could be adjusted according to the class of image one is searching for. In particular, the Least Square Projection (LSP) method [24], which has turned out to be a quite efficient MP method for CBIR [8], does not make use of any mechanism to tune similarity metrics so as to improve search results while querying multi-class image databases.

In this work we build on the flexibility provided by LSP and propose a new MP method, called *Class-Specific Multidimensional Projection* (CSMP), which not only preserves the good properties of LSP, but is also more accurate. In contrast to other MP methods; CSMP uses a family of metrics rather than a single one to accomplish the projection. More specifically, the CSMP combines pattern recognition tools such as feature extraction and feature selection to compute class-specific metrics tailored to maximizing the distances among instances belonging to different image classes. As shown in Section 4 (Experimental Results), the proposed class-specific metric construction increases accuracy considerably, outperforming other MP methods, including the original LSP method.

Paulo Joia · Erick Gomez-Nieto · João Batista Neto · Wallace Casaca
Glenda Botelho · Afonso Paiva · Luis Gustavo Nonato
Instituto de Ciências Matemáticas e de Computação – ICMC
Universidade de São Paulo – USP
São Carlos - SP – Brazil
E-mail: pjoia, egomezn, jbatista, wallace, glenda, apneto, gnonato@icmc.usp.br

Contributions We can summarize the main contributions of this work as:

- The design of class-specific metrics to measure the similarity between images (Section 3);
- A new multidimensional projection method (CSMP) that relies on the class-specific metrics to accomplish content-based image queries;
- A system capable of performing multiple queries in databases with images from many different classes;
- Unlike the previous MP approach [12], we do not employ query images as control points (control images). The initial projection and user interaction is carried out once over a small set of control images, leading to a final multidimensional projection. From this point on, any query image can be evaluated without recomputing the projections. This new strategy has turned out to be more efficient and generic than the previous one.

To the best of our knowledge, this is the first time an MP method makes use of a set of metrics rather than a single one to accomplish the mapping to the visual space.

2 Related work

As mentioned above, the technique presented in this paper makes use of multidimensional projection to perform content-based image retrieval. Therefore, we provide a brief overview of both fields in order to better contextualize our approach.

2.1 Content-Based Image Retrieval

Content-based Image Retrieval (CBIR) is a branch of Computer Science that encloses techniques and methods aimed at organizing large digital image repositories by means of visual content. As such, any technique ranging from a similarity function to a robust image annotation system are strictly related to CBIR [6].

Started in the 90's [30], research on CBIR has evolved to more specific topics such as relevance feedback [37], face recognition [36] and medical images [22]. Another comprehensive review has been written by *Datta et al.* [6].

In general, CBIR techniques seek to tackle two intrinsic problems: a) How can an image be mathematically described? and b) how to best compute the similarity between two images given their mathematical representation. The answers to these two questions are addressed in the following CBIR-related topics: feature extraction, feature selection, definition of distance functions and metrics, indexing of features and even user interfaces.

2.2 Multidimensional Projection

Multidimensional projection methods map instances from a high dimensional space into the visual space ($\mathbb{R}^2$ in our context) so as to preserve distances as much as possible. The MP methods can be grouped into two main categories, namely, global and local methods.

Global Methods map data instances into the visual space using a single transformation. Techniques based on spectral decomposition, also known as *classical scaling*, are typical examples of global methods. Classical scaling embeds instances from eigenvectors of a double-centered transformation applied to a dissimilarity matrix (symmetric matrix containing the dissimilarity between each pair of data instances) [34]. Although many authors have proposed alternatives towards getting around the high computational cost and the lack of flexibility during user interaction [2,3,9,27, 32,14], classical scaling is still computationally costly and cumbersome.

The global method proposed by Kruskal [15] uses non-linear-optimization to map data instances to the visual space. Since finding the minimum of the energy functional (commonly called stress function) is costly, Pekalska et al. [26] proposed a technique that first embeds a subset of samples in the visual space using Kruskal optimization scheme and then maps the remaining instances through a linear mapping. Although Pekalska's approach may be tuned to allow user interaction, it requires a minimum number of sample be projected in the first step, which hampers the user experience. The recent linear mapping PLMP [25] is similar to Pekalska's, but it makes use of faster mechanisms in both stages of the projection so as to enable the projection of out-of-core and streaming data.

Least Squares Projection [24] (LSP) is a two-step global technique that uses a non-linear scheme to position a subset of samples (control points) into the visual space. Assuming that each instance of data can be represented as an average of its neighbors, LSP builds Laplacian-like systems and uses the previously projected samples to constraint the system. Control points can be manipulated by user in order to facilitate grouping visualization and exploration.

Local Methods use, basically, two components to perform the multidimensional projection, namely the neighborhood information of each instance and the location of a subset of samples in the visual space. The mapping of an instance x is carried out by considering only the neighbors of x, which characterizes the local nature of the process. The well-known approach proposed by Chalmers [5] (and its hybrid variants [13,21]) first maps a subset of the samples using a force-based scheme and then exploits neighborhood structure of those samples to embed the remaining data in the visual space.

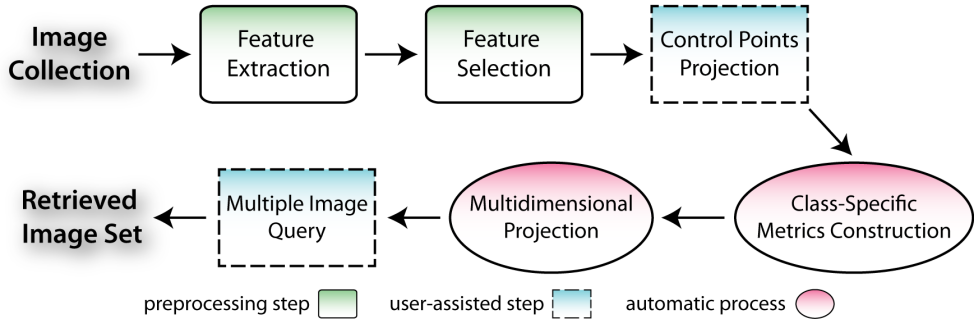


Fig. 1 Pipeline of our framework.

The PLP method [23] uses a force-based scheme to place a subset of samples in the visual space. The remaining data instances are projected through a family of Laplacian-like operators, which are built using local information. PLP provides great flexibility to user interaction, nonetheless, the continuous manipulation of projected data demands costly structural updates, impacting robustness and interactivity.

The recent LAMP method [11], relies on a mathematical formulation derived from orthogonal mapping theory. LAMP does not rely on neighborhood graphs and its mathematical formulation admits a quite reduced subset of sample instances as input. The orthogonality ensures that the resulting affine transformation behaves like a rigid transformation, that is, data can only be rotated and translated during the mapping process, avoiding scaling and shear effects. This behavior allows to preserve distances as much as possible during the multidimensional projection. The capability to deal with a quite reduces number of samples renders LAMP suitable for interactive applications.

The technique proposed in this work can be seen as a mid-term between global and local methods, since it uses a single transformation to project the instances onto the visual space while changing the metric to handle the data in a more localized manner. Therefore, our approach aims at holding the robustness of global methods while allowing for the flexible user interaction typically found in local methods.

3 CSMP Image Retrieval

The technicalities underlying the proposed projection-based image retrieval approach can be more conveniently described by the pipeline of Fig. 1. It consists of six main steps, namely, *feature extraction*, *feature selection*, *control points projection*, *class-specific metrics construction*, *multidimensional projection*, and *image query*. Both feature extraction and selection are typical non-interactive actions, carried out once for a given image dataset and, therefore, considered as preprocesses in our approach. User interaction takes place in both control points projection and image query steps. The remaining stages (class-specific metric construction and the multidimensional projection) are automatically computed, following every user-driven control point displacement. The following subsections go into more details on each of these processes.

3.1 Image Feature Extraction

Given a set of images $\mathscr{I}$, the very first step of our method corresponds to embed $\mathscr{I}$ in a feature space. Mathematically, we have to specify a transformation $\Lambda : \mathscr{I} \to \mathbb{R}^k$, where k is the number of features used to represent each image.

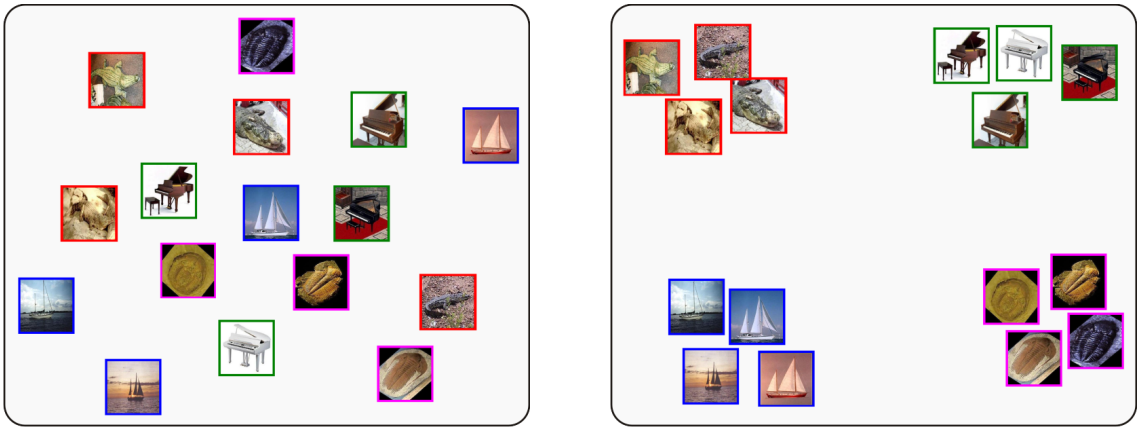
The transformation Λ is defined from feature extraction mechanism widely employed by the computer vision and pattern recognition communities. More specifically, we start by applying a sequence of feature extractors in each image $\alpha \in \mathscr{I}$, concatenating the results in a k-dimensional array. In particular, in our experiments we choose $k = 220$. The sequence of extractors we employ convey mainly texture (wavelet [17], Gabor [1], Tamura [7] and first order moment [33]) and color information (color moments [20]).

3.2 Image Feature Selection

Many of the features computed in the previous step may not be relevant to discriminate images. In order to clean up the feature array, we apply the well-known SSFS subset evaluator [10] tool which filters out irrelevant entries, retaining, in average, only 20% of the original amount of features, a much smaller feature subset.

Following this global feature selection process, we then identify the features that best characterize each class of a given pre-labeled image dataset. We used the Logistic Model Tree [19] as feature selector, since this method produced better results in most of the tests we conducted.

At the end of this process, we not only have a reduced feature vector, but also an indication of which of its features best discriminates each class of the image dataset, which is essential for the success of our technique. This selected subset will be then used in the remaining stages of our approach.



(a) (b)

Fig. 2 Control points (or images) projection: (a) Before user manipulation; (b) After user intervention, similar instances are grouped together.

3.3 Control Points Projection Stage

Suppose that $\mathscr{I}$ has already been embedded in the feature space $\mathbb{R}^k$. Our projection technique requires that a set of images (or points in projection context) is specified in order to drive the projection process. Such a set will be treated herein as *the control points set* or, simply, *control images*. In our context, this image set, denoted by Q, is provided by the user.

One strength of our method is that visual resources are available to the user during the process. Therefore, after selecting the control images the user can specify the class they belong to by interactively grouping similar control images in the visual space. Fig. 2(a) and 2(b), show, the resulting projection of the control images before and after user intervention, respectively.

3.4 Class-Specific Metrics Construction

Let Q_i be a subset of the control images that shares the same label (grouped together by the user) and $I_{Q_i} = \{i_1, \ldots, i_r\}$ be the indexes of the features that best represent the images in Q_i (determined by the feature selector). Given an image $\alpha \in \mathscr{I}$, $\beta \in Q_i$, and $\gamma \in Q_j$, $i \neq j$ it is reasonable to expect that

$$d_{Q_i}(\alpha,\beta) \leq d_{Q_j}(\alpha,\gamma)$$

if α belongs to the same class as β, where the *class-specific* metric d_{Q_i} is defined as

$$d_{Q_i}(\alpha,\beta) = \sum_{j \in I_{Q_i}} (\alpha^j - \beta^j)^2 \qquad (1)$$

where α^j accounts for the j^{th} coordinate (feature) of α (resp. β). The metric (1) is indeed a pseudo-metric, since $d_{Q_i}(x,y) = 0$ does not imply that $x = y$.

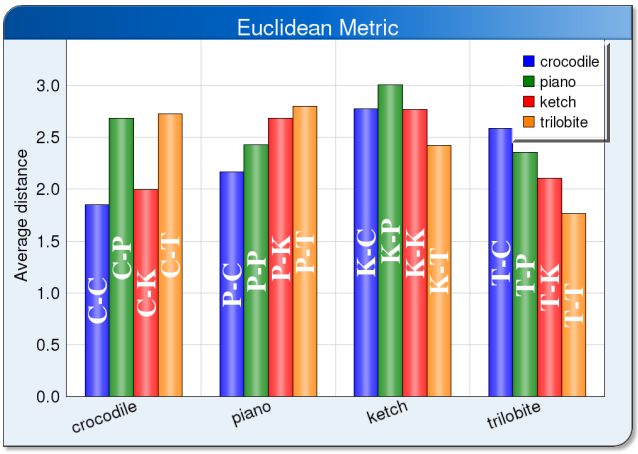
The rationale behind the class-specific distances defined in (1) is that if α is an image similar to $\beta \in Q_i$ then I_{Q_i} should be the best features to represent α, thus $d_{Q_i}(\alpha,\beta)$ should be small. However, one should expect a larger dissimilarity if a class-specific metric is used to measure the distance between α and a non-similar control image γ. Therefore, using class-specific metrics one avoids to compare features that do not represent the images properly, improving accuracy while increasing confidence in the value of the measured distance.

The bar graphs in Fig. 3 support our claim. Fig. 3(a) shows the average Euclidean distances for instances of four classes of images: “Crocodile”, “Piano”, “Ketch” and “Trilobite”. Notice that the distance between instances belonging to the same class such as Piano (P-P green bar) and Ketch (K-K red bar) are the same, or even greater, than the distance between instances belonging to different classes. This would definitely imply an inaccurate image retrievals. In contrast, as seen in Fig. 3(b), the class-specific distance measure brings all instances of a single class closer to their counterparts, while increasing the distance among instances belonging to other classes.

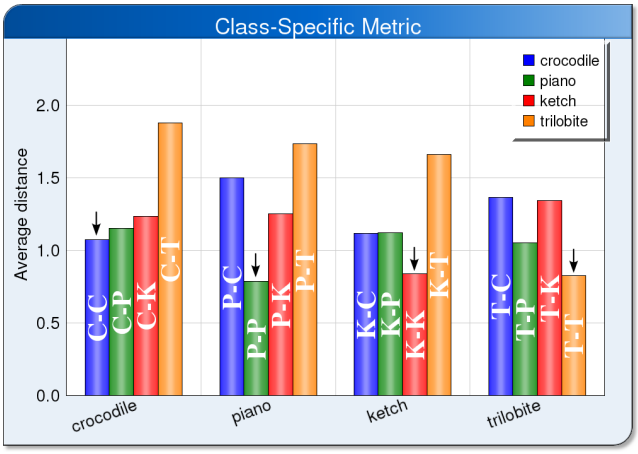
3.5 The Class-Specific Multidimensional Projection

The proposed Class-Specific Multidimensional Projection method builds on the LSP technique [24] to perform visual CBIR. However, in contrast to LSP, our approach makes use of the class-specific metrics defined in (1) so as to increase accuracy. Moreover, we constraint the linear system responsible for mapping instances to the visual space using a penalty method rather than the least square approach used by LSP. The penalty method has several advantages if compared to the least square method, as detailed below.

The CSMP relies on the assumption that each instance α of a dataset $\mathscr{I}$ can be written as a linear combination



(a) Average Euclidean distance



(b) Average class-specific distance

Fig. 3 Dataset containing 4 distinct classes of images. (a) Distances between classes measured using Euclidean metric. (b) Distances measured using class-specific metric. Notice that average distance between elements belonging to the same class (C-C, P-P, K-K, T-T) are smaller for class-specific distance (indicated by arrows).

of its nearest neighbors in the visual space. In more mathematical terms, let $N_\alpha = \{\alpha_1, \ldots, \alpha_s\}$ be the set of s nearest neighbors of α, and denote by (α_i^x, α_i^y) the coordinates of each element $\alpha_i \in N_\alpha$ when mapped to the visual space $\mathbb{R}^2$. From the linear combination hypothesis, one can compute the two-dimensional coordinates of α as:

$$(\alpha^x, \alpha^y) = \sum_{\alpha_i \in N_\alpha} c_{i\alpha}(\alpha_i^x, \alpha_i^y) \tag{2}$$

where $c_{i\alpha} > 0$.

Each image in $\mathscr{I}$ gives rise to a vectorial equation as the one given in (2), which can be assembled into two homogeneous linear systems:

$$L\mathbf{x} = 0; \quad L\mathbf{y} = 0 \tag{3}$$

where $\mathbf{x}$ and $\mathbf{y}$ account for the coordinates of the mapped elements and L is the matrix derived from equation (2).

The sets N_α define a *Nearest Neighbors Graph (NNG)* of $\mathscr{I}$, that is, a graph connecting each element in $\mathscr{I}$ to its nearest neighbors. It can be shown that the rank of L is $n-q$, where n is the number of elements in $\mathscr{I}$ and q is the number of connected components making up the NNG [31]. Thereby, in order to ensure a single non-trivial solution for the linear systems (3), the NNG should have only one connected component, which can be ensured by adding new edges linking disconnected components of the NNG.

The coefficients $c_{i\alpha}$ are defined as follows:

$$c_{i\alpha} = \begin{cases} d_{Q_i}(\alpha, \alpha_i) & \text{, if } \alpha \text{ or } \alpha_i \text{ is a control image} \\ d(\alpha, \alpha_i) & \text{, if } \alpha \text{ and } \alpha_i \text{ are not control images} \\ 0 & \text{, otherwise} \end{cases} \tag{4}$$

where d is the Euclidean distance and d_{Q_i} is the class-specific metric defined in (1). In order to ensure symmetry for L, we are assuming the convention $d_{Q_i}(\alpha, \alpha_i) = 0$, if α and α_i are control images from different classes.

The lack of geometric information in (3) may lead to solutions that are difficult to interpret and analyze. Geometrical information can be incorporated by using the coordinates of the user selected control images as constraints for (3).

We are using the penalty method [35] to constraint (3), which can be stated as follows: let Q be the control images and $\mathbf{b}^x$ (resp. $\mathbf{b}^y$) be the vector with zero in all entries except in the entries b_i corresponding to a control image α_i, where the value $b_i = \alpha_i^x$ is settled. The penalty method transforms the problem $L\mathbf{x} = 0$ into

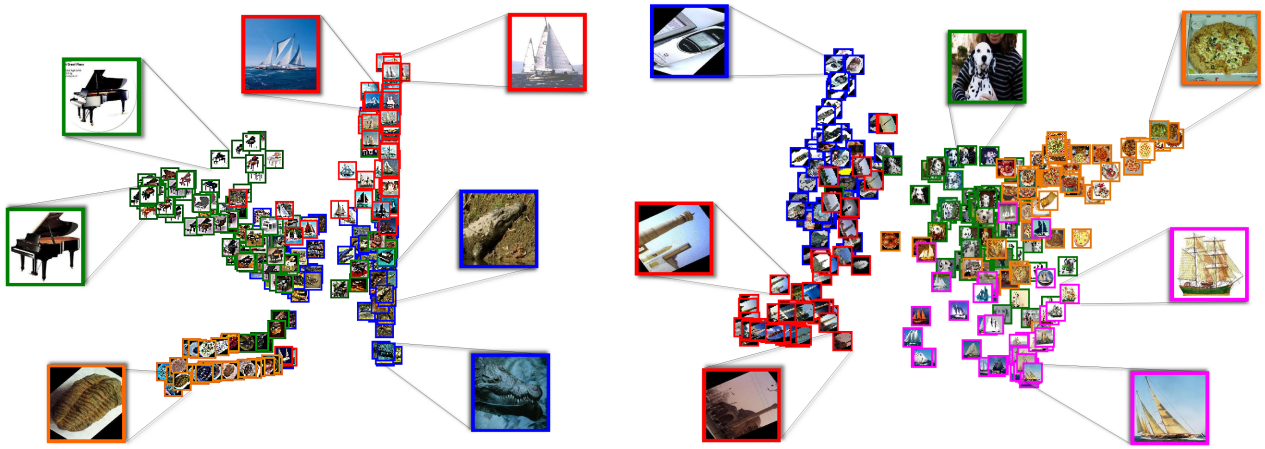
$$(L+P)f = Pb \tag{5}$$

where P is the diagonal penalty matrix with non-zero diagonal elements $p_{ii} = p$ only in the positions corresponding to the control images and p a large value (10^8 in our implementation).

The penalty method holds several good properties. For instance, it preserves the symmetry and positive semi-definiteness of the system, thus allowing for Cholesky factorization. Moreover, adding a large positive value in some diagonal entries improves the conditioning number of the matrix, thus avoiding the numerical instabilities.

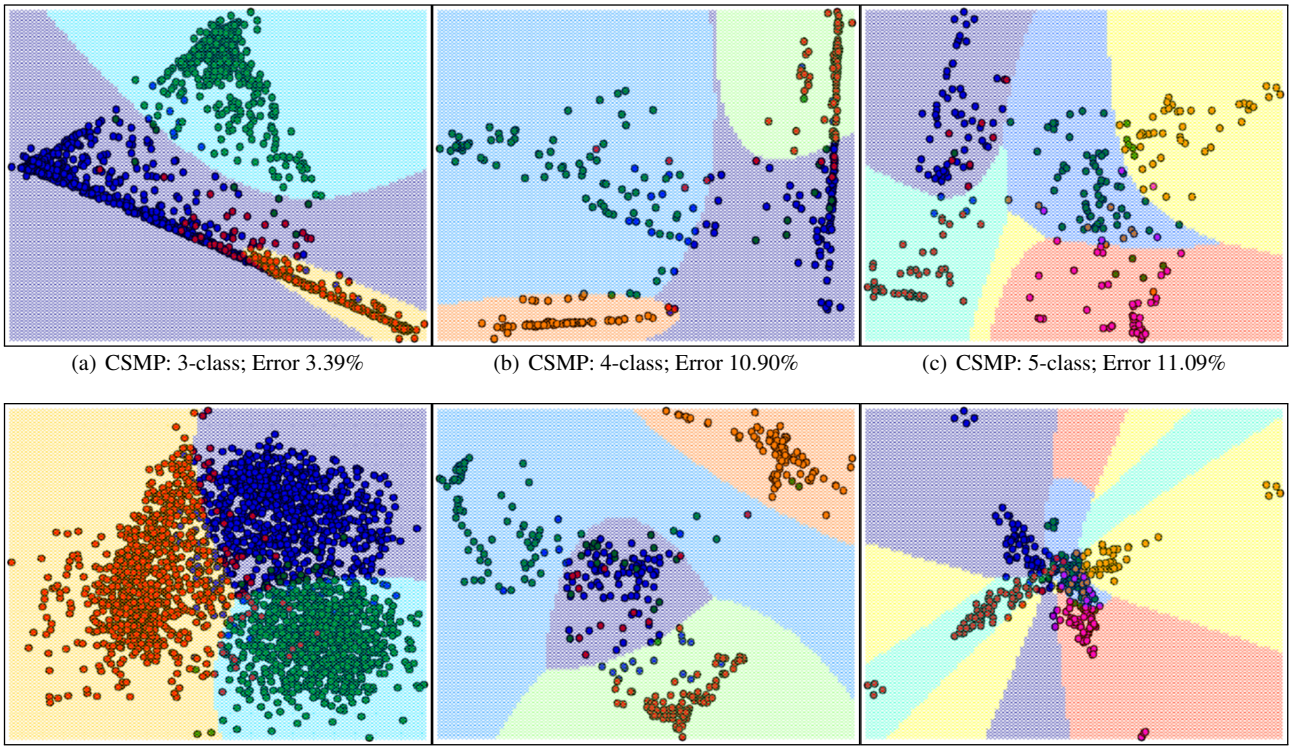
3.6 Multiple Image Query : a CBIR Approach

It is important to point out that unlike our original proposal [12], we compute, from the final multidimensional projection, the distances among all vertices. That is, all projected points (images) are modeled as a complete graph, yielding a total of $(n^2-n)/2$ edges, or distance measures. By storing this measures in ascending order, the retrieval of the m most similar images for a given query image becomes trivial. Moreover, given the low cost of this procedure, we can do that simultaneously, i.e., for as many query images as the user selects.



(a) 4-class dataset: distinct color frames depict distinct classes (b) 5-class dataset: distinct color frames depict distinct classes

Fig. 4 Projection of the thumbnails obtained by the CSMP projections for the 4-class and 5-class data sets shown in Figs. 5(b) and 5(c), respectively.



(a) CSMP: 3-class; Error 3.39% (b) CSMP: 4-class; Error 10.90% (c) CSMP: 5-class; Error 11.09%

(d) PLMP: 3-class; Error 7.61% (e) LSP: 4-class; Error 14.13% (f) LAMP: 5-class; Error 16.01%

Fig. 5 Comparing quality of projection. (a) and (d): CSMP (proposed method) x PLMP for the 3-class dataset; (b) and (e): CSMP x LSP for the 4-class dataset; (c) and (f): CSMP x LAMP for the 5-class dataset.

4 Experimental Results

To evaluate the proposed approach, two experiments are provided. The first one aims to show how the projections produced by the CSMP compare with those computed by LAMP, LSP and PLMP counterparts. Both qualitative and quantitative results are given. The second experiment describes the CSMP in a CBIR context.

Table 1 gives some details on the three data sets used, taken from the Caltech101 [18] 256x256 image database. Figs. 5(a), (b) and (c) show the projections with CSMP for the 3-, 4- and 5-class data sets, respectively.

Table 1 Data sets used in the experiments, from left to right the columns correspond to the dataset name, classes, size and dimension.

Name	Classes	Size	Dim*
caltech-3class	airplane(800), faces(870) and motorbikes(798)	2468	55
caltech-4class	crocodile(101), ketch(114), grand-piano(99) and trilobite(86)	400	48
caltech-5class	cellphone(59), dalmatian(67), minaret(76), pizza(53) and schooner(63)	318	59

* Dimension after features selection.

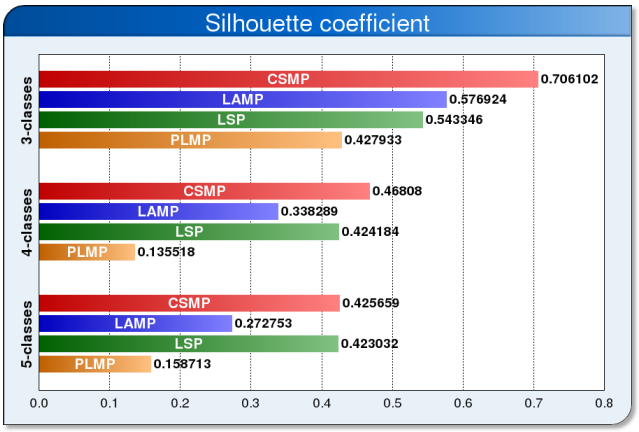


Fig. 6 Silhouette coefficient for every used dataset: 3, 4, 5-class. CSMP projection (proposed method) is better than the other three projection techniques, for all data sets.

4.1 Experiment 1 - Assessing the Quality of CSMP Projections

In this experiment we compare the CSMP projections, for all three classe, with LAMP, LSP and PLMP. Results are illustrated in Fig. 4 and Fig. 5. Notice that CSMP error percentage is smaller than PLMP (Fig. 5(a),(d)), LSP (Fig. 5(b),(e)) and LAMP (Fig. 5(c),(f)).

To estimate the separability among classes, the visual space is divided into regions, according to a discriminant function (Discriminant Function Analysis). In this experiment we opted for a quadratic function to capture the largest possible number of elements of a given class within a region. This function determines which elements of a dataset may be considered as belonging to the same class by means of some statistical computation [16,28]. Regions and projections can then be visualized as shown in Fig. 5. It can be seen that the CSMP produced the lowest error rates for all data sets in comparison with other techniques. Moreover, the data is much better grouped when CSMP is used, which is an important aspect if data is not labeled, since clusters have to be visually identified.

Projections such as those of Fig. 5 can also be quantitatively evaluated by silhouette coefficient [4]. This coefficient measures both the cohesion and the separation between grouped instances. The cohesion a_x of x is calculated as the average of the distances between x and all other instances belonging to the same group as x. The separation b_x is the minimum distance between x and all other instances belonging to other groups. The silhouette of a projection is given by:

$$Silh = \frac{1}{n} \sum_{x \in \mathscr{X}} \frac{(b_x - a_x)}{max(a_x, b_x)},$$

where n is the number of instances. It is a sensitive coefficient in the interval $[-1,1]$. The higher the values of silhouette for a given projection, the better the cohesion and the separation, that is, instances belonging to the same class are closer to each other, and yet, distinct classes are farther apart. Hence, higher values of silhouette indicate better projections.

The bar graphs in Fig. 6 show the silhouette values computed from projections. The CSMP silhouettes are higher than those for LAMP, LSP and PLMP approaches, for all data sets. It should be remembered that even small variations in silhouette (third or fourth decimal place) can result in better data clustering, as this coefficient is capable of capturing very small changes in the clustering map.

4.2 Experiment 2 - CSMP in a CBIR Context

In order to further attest the effectiveness of our approach we compare CSMP against two public CBIR systems: FIRE [7] and GA [29], both for the 5-class dataset. Fig.7 shows the resulting error of the CBIR queries for both methods. The errors rates in the y-axis were computed as follows: for each set of retrieved images $(5,10,\cdots,30)$, we performed three queries, for three distinct images randomly selected.

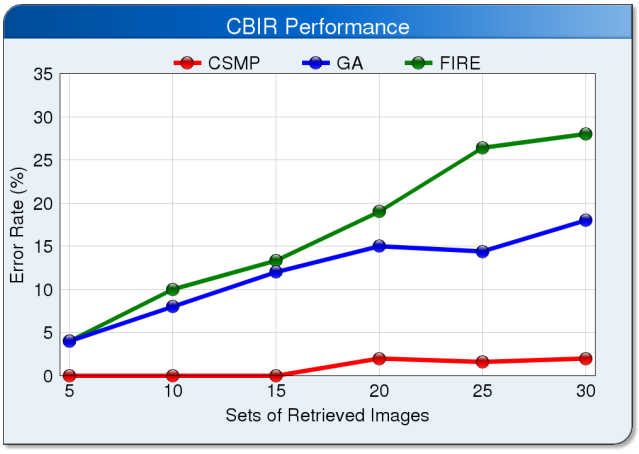
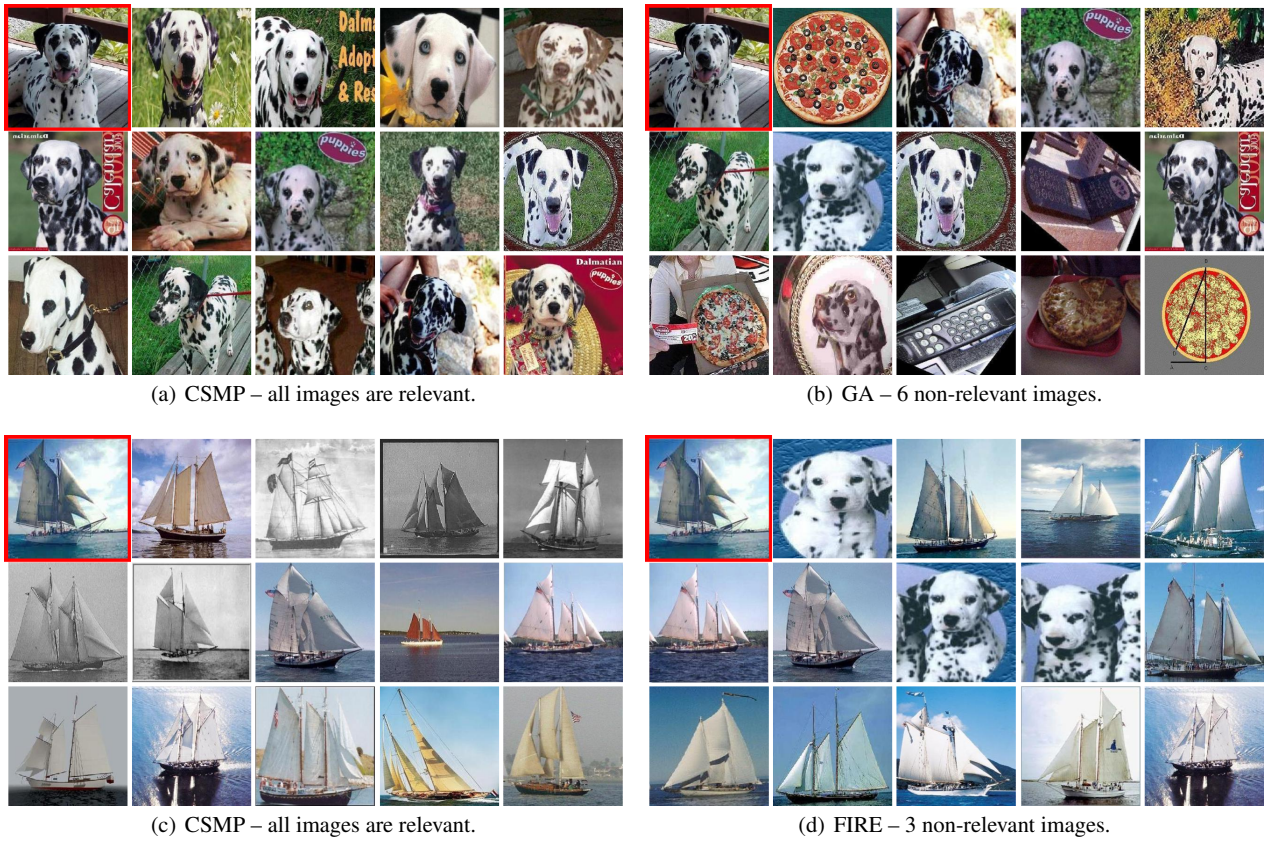


Fig. 7 CSMP against other CBIR systems for the 5-class dataset.



(a) CSMP – all images are relevant.

(b) GA – 6 non-relevant images.

(c) CSMP – all images are relevant.

(d) FIRE – 3 non-relevant images.

Fig. 9 Image retrieval for CSMP, GA and FIRE with top 15 images. In red frame, the image used as query. Notice that all images in CSMP are relevant, whereas GA and FIRE retrieved six and three non-relevant images, respectively.

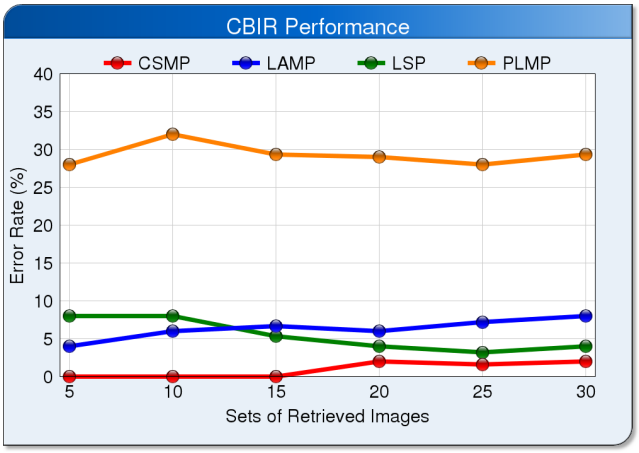


Fig. 8 CSMP against other projection techniques in CBIR context, for the 5-class dataset.

We then computed the mean value for the three queries based on the relevant and non-relevant images retrieved. It can be seen that our method produces lower error rates in all retrievals.

A similar comparison was conducted with the projection techniques: LAMP, LSP and PLMP. Results can be seen in Fig. 8. Notice that CSMP also presented the lowest error rates. However, projection techniques were more competitive than typical CBIR systems, mainly the classical LSP technique.

Visual comparison between CSMP and both CBIR systems is presented in Fig. 9, for the 5-class dataset. The first thumbnail, in red frame, is the query image used. The rank list of images is displayed in descending similarity order. Figs. 9(a) and (b) show, respectively, the outputs for CSMP and GA CBIR system. Observe that CSMP outperforms the GA CBIR considerably, since all images displayed are relevant and belong to the class of images *dalmatian* (Fig. 9(a)), whereas six non-relevant images were retrieved by GA CBIR (Fig. 9(b)). Figs. 9(c) and (d) show, respectively, results for CSMP and FIRE CBIR. Again CSMP is superior: all images returned by the former are relevant (Fig. 9(c)), whereas three non-relevant images appear in the latter (Fig. 9(d)).

Fig. 10 shows the visual results between CSMP and traditional CBIR systems for multiple image queries in the 5-class dataset. In contrast with the GA and FIRE CBIR systems, CSMP allows multiple image queries, keeping the correlation between different classes of images without recomputing the multidimensional projection multiple times. Meanwhile, to perform multiple image queries in GA and

(a) CSMP – all images are relevant

(b) GA – 2 non-relevant images, both in class cellphone.

(c) FIRE – 3 non-relevant images, 2 in class pizza and 1 in class cellphone.

Fig. 10 Multiple image queries by image retrieval using CSMP, GA and FIRE with top 30 images from two image queries. In red frame, the images used as queries. Notice that all images in CSMP are relevant, whereas GA and FIRE retrieved two and three non-relevant images, respectively.

FIRE CBIR systems, we compute two single image queries independently, one for each query image. Then, we combine the outputs of the single queries to produce the final result. The thumbnails in red frames are the query images used and the similarity of retrieved images increasing from right to left. Figs. 10(a), (b) and (c) show, respectively, the outputs for CSMP, GA and FIRE CBIR. CSMP has a better performance than GA and FIRE CBIR, since all retrieved images are relevant and belong to the classes of images *pizza* and *cellphone* from Caltech101 image collection.

5 Conclusion and Future Work

The results presented in Section 4 clearly show the effectiveness of the CSMP technique, surpassing, in both accuracy and flexibility, state-of-the-art methods. The superior performance of CSMP is a consequence of the proposed class-specific distance measure it relies on, which ensures a more reliable distance definition among similar instances. Simplicity is another advantage of CSMP, which essentially requires a linear solver library to be implemented.

The ability to perform multiple queries is another important property of CSMP that many applications can benefit from. This property allied to the possibility of interactively changing the position of control points in the visual space render CSMP a very attractive CBIR method.

In our experiments we notice that more "spread" layouts are produced when penalty factor p (set equal to 10^8 in all results presented in this paper) is decreased. Choosing the ideal number of neighbors of each instance is another aspect we have to investigate more deeply. An alternative to the k-nearest neighbors scheme employed in our implementation would be to define a radius of influence to each point. However, finding the appropriate radius is not an easy task either.

Acknowledgment

We thank the anonymous reviewers for their useful and constructive comments. This work was supported by FAPESP and CAPES-Brazil.

References

1. Arivazhagan, S., Ganesan, L., Priyal, S.P.: Texture classification using gabor wavelets based rotation invariant features. Pattern Recogn. Lett. **27**, 1976–1982 (2006)
2. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput. **15**(6), 1373–1396 (2003)
3. Brandes, U., Pich, C.: Eigensolver methods for progressive multidimensional scaling of large data. pp. 42–53 (2007)
4. Campello, R.J., Hruschka, E.R., Alves, V.S.: On the efficiency of evolutionary fuzzy clustering. J. Heuristics **15**, 43–75 (2009)

5. Chalmers, M.: A linear iteration time layout algorithm for visualising high-dimensional data. In: Proc. of Visualization '96, pp. 127 –131 (1996)
6. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image retrieval: Ideas, influences, and trends of the new age. ACM Comput. Surv. **40**, 1–60 (2008)
7. Deselaers, T., Keysers, D., Ney, H.: Features for image retrieval: An experimental comparison. Inform. Retrieval **11**, 77–107 (2008)
8. Eler, D., Nakazaki, M., Paulovich, F.V., Santos, D., Andery, G., Oliveira, M., Batista Neto, J., Minghim, R.: Visual analysis of image collections. Visual Comput. **25**, 923–937 (2009)
9. Faloutsos, C., Lin, K.: Fastmap: a fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. Proc. of SIGMOD'95 **24**, 163–174 (1995)
10. Gutlein, M., Frank, E., Hall, M., Karwath, A.: Large-scale attribute selection using wrappers. In: Proc. of IEEE CIDM'09, pp. 332–339 (2009)
11. Joia, P., Coimbra, D., Cuminato, J.A., Paulovich, F.V., Nonato, L.G.: Local Affine Multidimensional Projection. IEEE T. on Vis. and Comput. Gr. **17**, 2563–2571 (2011)
12. Joia, P., Gomez-Nieto, E., Botelho, G., Batista Neto, J., Paiva, A., Nonato, L.G.: Projection-based image retrieval using class-specific metrics. In: Proc. of SIBGRAPI 2011 (2011)
13. Jourdan, F., Melangon, G.: Multiscale hybrid MDS. In: Proc. of IEEE InfoVis 2004, pp. 388–393 (2004)
14. Koren, Y., Carmel, L., Harel, D.: Ace: a fast multiscale eigenvectors computation for drawing huge graphs. In: Proc. of IEEE InfoVis 2002, pp. 137–144 (2002)
15. Kruskal, J.: Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika **29**, 1–27 (1964)
16. Krzanowski, W.J.: Principles of Multivariate Analysis: a user's perspective. Oxford (1988)
17. Kumar, D.A., Esther, J.: Comparative study on CBIR based by color histogram, Gabor and wavelet transform. Int. J. Comput. Appl. **17**(3), 37–44 (2011)
18. L. Fei-Fei, R.F., Perona, P.: Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories (2004)
19. Landwehr, N., Hall, M., Frank, E.: Logistic model trees. Mach. Learn. **59**, 161–205 (2005)
20. Maheshwary, P., Srivastav, N.: Retrieving similar image using color moment feature detector and k-means clustering of remote sensing images. In: Proc. of ICCEE 2008, pp. 821–824 (2008)
21. Morrison, A., Ross, G., Chalmers, M.: A hybrid layout algorithm for sub-quadratic multidimensional scaling. In: Proc. of IEEE InfoVis 2002, pp. 152–158 (2002)
22. Muller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of content-based image retrieval systems in medical applications – clinical benefits and future directions. Int. J. of M. Inform. **73**(1), 1–23 (2004)
23. Paulovich, F.V., Eler, D.M., Poco, J., Botha, C.P., Minghim, R., Nonato, L.G.: Piecewise laplacian-based projection for interactive data exploration and organization. Comput. Graph. Forum **30**, 1091–1100 (2011)
24. Paulovich, F.V., Nonato, L.G., Minghim, R., Levkowitz, H.: Least square projection: A fast high-precision multidimensional projection technique and its application to document mapping. IEEE T. on Vis. and Comput. Gr. **14**(3), 564–575 (2008)
25. Paulovich, F.V., Silva, C., Nonato, L.G.: Two-phase mapping for projecting massive data sets. IEEE T. on Vis. and Comput. Gr. **16**(6), 1281–1290 (2010)
26. Pekalska, E., de Ridder, D., Duin, R.P.W., Kraaijweld, M.A.: A new method of generalizing Sammon mapping with application to algorithm speed-up. In: Annual Conf. Adv. School for Comput. Imag., pp. 221–228 (1999)
27. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science **290**(5500), 2323–2326 (2000)
28. Seber, G.A.F.: Multivariate Observations. Wiley (1984)
29. da Silva, S.F., Batista, M.A., Barcelos, C.A.Z.: Adaptive image retrieval through the use of a genetic algorithm. In: Proc. of IEEE ICTAI 2007, vol. 1, pp. 557–564 (2007)
30. Smeulders, A., Worring, M., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE T. Pattern Anal. Mach. Intell. **22**(12), 1349–1380 (2000)
31. Sorkine, O., Cohen-Or, D., Lipman, Y., Alexa, M., Rössl, C., H.-P.Seidel: Laplacian surface editing. In: Proc. of SGP 2004, pp. 175–184 (2004)
32. Tenenbaum, J.B., Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science **290**(5500), 2319–2323 (2000)
33. Theodoridis, S., Koutroumbas, K.: Pattern Recognition. Academic Press (2009)
34. Torgerson, W.: Multidimensional scaling of similarity. Psychometrika **30**, 379–393 (1965)
35. Xu, K., Zhang, H., Cohen-Or, D., Xiong, Y.: Dynamic harmonic fields for surface processing. Comput. Graph. **33**(3), 391–398 (2009)
36. Zhao, W., Chellapa, R., Phillips, P., Rosenfeld, A.: Face recognition: A literature survey. ACM Comput. Surv. **8**(4), 399–458 (2000)
37. Zhou, X., Huang, T.: Relevance feedback in image retrieval: A comprehensive review. Multimidia Syst. **40**, 262–282 (2003)