Visual Data Exploration to Feature Space Definition

Bruno Brandoli, Danilo Eler, Fernando Paulovich, Rosane Minghim, Joao Batista
Instituto de Ciências Matemáticas e de Computação (ICMC) – University of São Paulo (USP)
São Carlos, Brazil
Email: {brandoli,eler,paulovic,minghim,jbatista}@icmc.usp.br

***Abstract*—Many image-related applications rely on the fact that the dataset under investigation is correctly represented by features. However, defining a set of features that properly represents a dataset is still a challenging and, in most cases, an exhausting task. Most of the available techniques, especially when a large number of features is considered, are based on purely quantitative statistical measures or approaches based on artificial intelligence, and normally are "black-boxes" to the user. The approach proposed here seeks to open this "black-box" by means of visual representations, enabling users to get insight about the meaning and representativeness of the features computed from different feature extraction algorithms and sets of parameters. The results show that, as the combination of sets of features and changes in parameters improves the quality of the visual representation, the accuracy of the classification for the computed features also improves. The results strongly suggest that our approach can be successfully employed as a guidance to defining and understanding a set of features that properly represents an image dataset.**

***Keywords*-Visual Feature Space Analysis; Feature Space Visualization; Feature Space Evaluation; Visual Exploration**

I. Introduction

The extraction of feature vectors is crucial for many image-related applications, for instance *pattern recognition*, *content-based image retrieval* or *image mining*, determining the accuracy of the results achieved by them [1]. There exist several algorithms to compute such vectors, each of which with their own set of parameters, reflecting different properties of the image dataset under investigation. Therefore, the choice of the set of features that provides the highest classification rates or the most efficient retrieval poses as a challenging accomplishment. Usually, such task requires *a priori* knowledge of the image dataset [2]. Also, to build a robust model several experiments may be required and this is often a very expensive process [3].

In many cases, users start by selecting a pre-labeled set of images and define the parameters for the feature extraction algorithm, which is frequently a tedious and laborious task. Then, feature vectors are computed and classification attained. High classification rates will be only achieved if the set of features proves adequate for the image dataset. This can only be assessed after the classification process has been carried out. In other words, one cannot infer beforehand how effective the computed features will be for the classification process. Our goal is to provide effective tools to support feature definition speeding up the setup procedure before classification.

In this paper we propose a new approach to the visual analysis of feature spaces using point placement or projection techniques [4]. Projection techniques work by mapping high-dimensional data into a lower dimensional visual space, whilst retaining, up to an extent, the distance relationships defined in the original data space. Initially, a set of features is extracted. Then, the computed feature vectors are visualized in a 2D representation that reveals the similarity relationships between the images under analysis. This visual representation is used to determine if this set of features successfully represents the image dataset, according to an expert point of view. If the similarity relationships match with what is expected by the expert – similar images are closely placed and the dissimilar ones are positioned far apart –, the set of features properly represents the dataset and can be considered for other tasks, such as classification. Otherwise, the parameters can be changed or another extraction algorithm can be employed, producing a new set of features which can then be visually explored to check if it properly represents the dataset. This provides an interactive visual approach which help users to construct better models for image processing.

The main contributions of this paper are:

- an approach for visual exploration of feature spaces targeted at converging to useful features for image processing, according to the expert point of view;
- an interactive visual framework which help users to better "understand" different sets of features;
- a method to objectively evaluate the quality of projections which match with the concept that better projections result in more appropriate feature spaces, specially for classification tasks.

The remaining of this paper is organized as follows. Section II describes related work on visual analysis of feature spaces. Section III presents background information on feature extraction and information visualization techniques employed here. Section IV details our approach through a simple example. Section V presents the results of the experiments performed in order to evaluate the proposed approach. Finally, conclusions and directions for future research are given in Section VI.

II. Related Work

Employing visualization techniques to explore and to draw knowledge from datasets is an efficient way to combine human intelligence with the powerful force of computation [5]. Several visualization techniques and tools have been developed enabling users to interact with abstract data [6], some of which specially designed to explore multidimensional spaces resulting from the feature extraction process.

One of these approaches was proposed by Rodrigues et. al. [7] (further extended in [8]). The aim is to give support to the analysis of the features employed on similarity queries for a content-based image retrieval system. Once the feature space is composed, a visual representation is created showing that the best visual representation conveys the best precision and recall measures when a query is executed on the data. However, the visual representation is not used to help an user to interactively define the best set of features or the best set of parameters to extract the features, but only to confirm that the query precision and recall matches with the visual representation quality. In our approach, the visual representation is meant as an interactive guide to explore, define and refine the set of features, giving insight on how the extraction algorithm parameters or the weighting of the features affect the similarity relationships between groups or individual images.

PEx-Image [9] is a similar tool, which employs a point placement visual representation to explore image collections. This tool provides interactive visualizations to aid on the exploration of feature spaces, and supports the comparison between different spaces using coordination techniques. The main difference between PEx-Image and the approach proposed is that the former aims at creating the best visual representation given a feature space, while here we seek to create a visual representation which best reflects the feature space, that is, a visual representation that is as good as the feature space. Therefore, our approach gives more precise insight in the similarity relationships between images. In addition, in PEx-Image the quality of a visual representation is only defined according to the user's point of view. Here we use a well-known measure, borrowed from the data-mining community, to help users assess such quality, thus reducing the subjectiveness of conclusions based purely on the visual analysis.

III. Background

A. Feature Extraction

Feature extraction is the process of capturing quantitative characteristics of an $\Re^{w \times h}$ image and place them into a $\Re^n$ dimensional feature vector, in which n is the number of values extracted, w and h are image width and height, respectively. Several feature extraction methods, also known as descriptors, have been proposed in the literature [10]. In this paper, traditional texture analysis methods as co-occurrence matrix [11] and Gabor filters [12] features are used. In the former, 5 measures (*energy*, *entropy*, *inertia*, *inverse difference moment*, and *correlation*), 5 distances and 4 directions are considered, summing-up 100 distinct features. In the latter, 16 features using energy of the responses of Gabor filters are computed (4 orientations - 0^0, 45^0, 90^0 and 135^0 - and 4 scales). In addition, experiments are also carried out using features extracted from R, G, B plane, yielding 48 features.

We also test features extract using the *bag-of-visual-features (BoVF)* model. The process for constructing a BoVF model starts by selecting a set of keypoints over the images. This selection is done sparsely or densely. Several local regions detectors have been proposed in the literature [13]. Then, the descriptors are computed around the keypoints. In this work, the keypoints are obtained with the Harris-Laplace point detector or dense sampling [13]. After the points selection and description, each vector is quantized against a visual codebook or vocabulary. Codebooks are usually constructed by using a method to cluster the keypoints. In this work, the visual codebook is created via *k-means* [14] clustering algorithm. Once the vocabulary is defined, each keypoint is assigned to the codebook element that is closest in the Euclidean space. The result is a histogram that represents the BoVF model normalized to sum-up 1.

In this work, a visual representation of the feature space is carried out right after feature extraction. Such visual representations are constructed by employing *multidimensional projections techniques*, which are detailed in the next section.

B. Multidimensional Projection Techniques

Multidimensional projection techniques, or simply *projection techniques*, seek to create visual representations that enable users to employ their visual ability to recognize patterns and structures present in the dataset. Each data instance (an image) is represented as a visual element, such as a circle, point or sphere, and mapped into a visual space that may be either 1D, 2D, or 3D. The relative positions of these elements reflect some type of relationship between data instances, the most common being the similarities or neighborhood relationships [15]. In this case, if the elements are closely placed on the final layout, it indicates that the data instances they represent are similar according to a certain distance. If the elements are projected far apart, it indicates that the objects they represent are not related.

Currently, there exist a number of different techniques considering different aspects of data distribution, such as the distance distribution, non-linear relationships among the dimensions, etc. (we refer to [15] for more details). Here we are interested in a very accurate technique, that is, the one which reflects as much as possible on the visual layout the distance relationships between the data instances in the original space. To accomplish that *Classical Scaling* is considered the best choice. In *Classical Scaling*, a doublycentered distance matrix between all pairs of data instances is defined, and a spectral decomposition is applied to recover the Cartesian coordinates of the elements in the visual space. It is possible to prove that if the distance function is Euclidean, the projected space presents the smallest mean quadratic deviation from the original space between all possible reduced spaces. Therefore, preserving as much as possible the distance relationships amongst the data instances when they are projected on the visual space [16]. Formally, d_{ij} is the distance between points x_i and x_j, and placed in the position r_i and r_j in projection, respectively. Finally, the minimum of the Equation 1 is computed.

$$P = \sum_{i} \sum_{j>i} \frac{(d_{ij} - \|r_i - r_j\|)^2}{d_{ij}} \tag{1}$$

C. Assessing the Projection Quality

The visual analysis approach proposed in this work aims at verifying whether similar images, according to the user's point of view, are also similar according to the extracted set of features used to represent them. Although users may employ their visual abilities to determine the quality of a projection, it is rather difficult to tell small differences resulted from slightly modified set of features. In order to reduce this subjectiveness, a measure called *silhouette coefficient* [17], which was originally proposed to evaluate results of clustering algorithms, is employed.

The silhouette coefficient measures both the cohesion and separation between instances of clusters. Consider an instance i belonging to cluster, its cohesion a_i is calculated as the average of the distances between i and all other instances belonging to the same cluster. The separation b_i is the minimum distance between i and all other instances belonging to the other clusters. The silhouette of a projection is given as the average of the silhouette of all instances, where n is the number of instances. In Equation (2) it is formalized.

$$S = \frac{1}{n} \sum_{i=1}^{n} \frac{(b_i - a_i)}{\max(a_i, b_i)} \tag{2}$$

The silhouette can vary between $-1 \leq S \leq 1$. Larger values indicate better cohesion and separation between clusters. In our approach, clusters are composed taking into account the pre-labeled instances, and the silhouette indicates whether images belonging to the same class are more similar between themselves than images belonging to other classes.

The usefulness of our approach is presented in the next section by means of a simple application.

IV. An Approach to Visual Feature Space Analysis

The diagram of Figure 1 summarizes the approach suggested here. First, a set of image features is extracted and then projected onto a 2D space. Upon user's agreement on the goodness of the provided result (based on visual inspection or observation of the silhouette coefficient), the set of feature can then be used for other tasks, such as classification, retrieval or data mining. If significant misplacement is found, the parameters or extraction methods used to compute features ought to be changed, and the process re-executed. This process is performed until the expected outcome is reached, that is, until the projection separates what is similar from what is not.

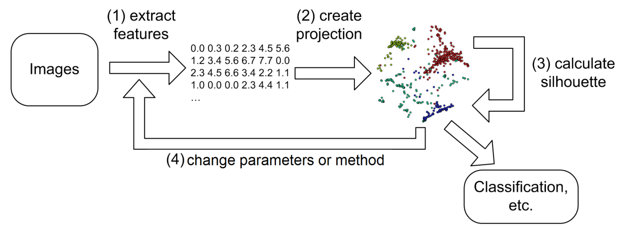


Figure 1. Proposed visual exploration of the feature space.

To exemplify our approach we propose an initial experiment formed by a set of 70 texture images from the Brodatz dataset [18], grouped into 7 classes with 10 images each. Features are extracted using Gabor filters. The feature space visualization is depicted in Figure 2. Images are represented as circles, and are colored according to the class they belong to. For clarity, one image sample of each class is placed near to its corresponding cluster. In this projection 5 classes are well separated, but 2 are mixed up (top right-hand corner). This is a clear indication of a limited discriminant power of the chosen feature set. A closer look, on the other hand, reveals that in terms of texture pattern, both mixed up classes are very similar. Therefore, it is up to the user to decide whether such classes should in fact be considered as a single class. If that is the case, this can be taken as a good set of features. Otherwise, a new set should be evaluated.

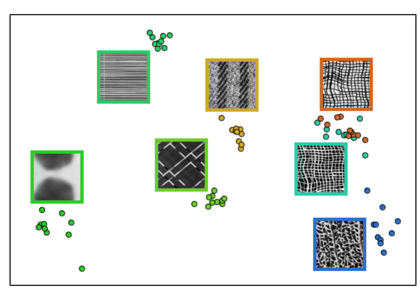
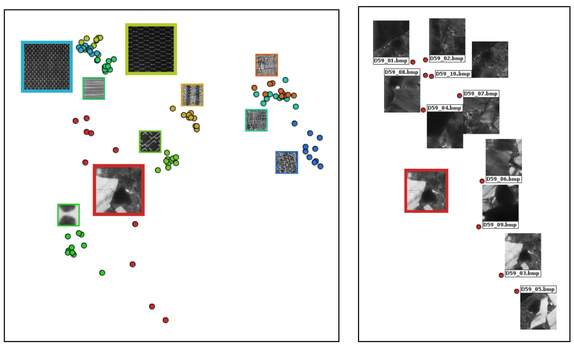


Figure 2. Analysis of an image dataset containing 7 different classes. 5 classes of images are well distinguished, but 2 are mixed-up, indicating that the employed features could not separate these 2 classes.

The experiment is extended with the addition of 3 new texture classes (marble and two wired-frame classes), and the resulting projection is shown in Figure 3(a). The same set of parameters is used. Observe that the marble class, whose sample is highlighted in red, appears scattered throughout the projection. The marble images normally exhibit non-uniform texture elements, which vary both in size and orientation. This behavior is illustrated in Figure 3(b) that shows the zoomed-in area of the marble class depicted in Figure 3(a). Since a fixed set of Gabor filters parameters (for orientation and scales) is used, it is unlikely that all subtleties in marble textures with varied sizes and orientation can possibly be captured.



(a) (b)

Figure 3. Projection of 10 classes image dataset (a) and a zoom in of the marble class (b). The employed features cannot distinguish the different wire-frame patterns, and do not define a uniform marble class.

For the wired-frame classes, the projected samples are mixed up, as can be observed at the top left-hand corner of Figure 3(a). To reach better separability, only Gabor features with 90^0 orientation are now selected and a new projection is computed, as shown in Figure 4. Notice that these two classes are now separated, for texture patterns with 90^0 orientation are different in each class. However, when compared to the previous projection, the remaining classes are more spread. In this case, the user can conclude that the orientation 90^0 is effective in separating these 2 new wired-frame classes of images, but the cohesion of the other classes is impaired. This conclusion is not easy to reach without feature space visualization. The silhouette coefficients computed for Figure 4 and Figure 3(a) are 0.429 and 0.474, respectively. Although similar, those values inidicate that the features explored in Figure 4 are better than those explored in Figure 3(a), matching with the visual inspection.

Whichever orientations have been chosen in the previous experiment, it is possible to notice that Gabor features failed to separate samples of the two Brodatz's Linen classes. Their samples are colored in brown and cyan (top-right corner of Figure 2). They are also present in Figures 3(a) and 4. These two classes exhibit very similar texture patterns and closer inspection reveals a slight variation in the pixel intensities only. Seeking better separability, we investigate new features produced by the co-occurrence matrix technique, parameterized as described in Section III. The resulting projection is shown in Figure 5. Notice that the brown

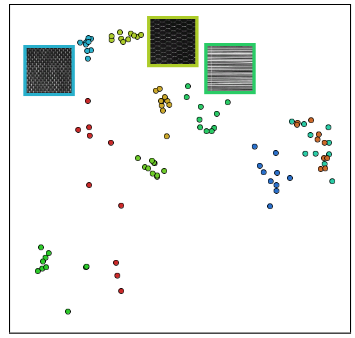


Figure 4. Projection for the same dataset shown in Figure 3, but with a small set of Gabor features. Now the 2 wired-frame classes are better separated, although the cohesion of other classes is impaired.

and cyan samples (enlarged icons) are now well separated, as well as the other classes of images, indicating that the co-occurrence features, when compared with Gabor filters, produce a better representation for the dataset. The silhouette coefficient reinforces this perception. Projection with co-occurrence features is 0.583, against 0.474 for the Gabor experiment.

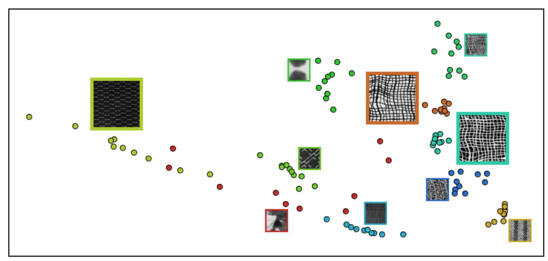


Figure 5. Projection for the same dataset shown in Figure 4 with co-occurrence features. Both visual inspection and higher silhouette coefficient indicate a better representation for the dataset if compared to that shown in Figure 3.

In the next section we apply the same approach to more complex datasets, showing that it works in different scenarios.

V. Experimental Results

In order to evaluate our approach, empirical experiments are executed on three different image datasets. Next they are described.

A. Datasets

The KTH-TIPS database [19] consists of 10 colorful texture classes. Images are taken at 9 different scales. Each scale is viewed under three illumination directions and three poses, resulting a total of 9 images per scale, and 81 images per material.

The Corel-1000 database [20] contains 1000 images from natural scenes and artificial objects separated in 10 classes.

It is frequently used for the evaluation of image retrieval methods.

The Caltech database [21] contains 6 unbalanced classes and a total of $3,812$ images: airplane_side (1074 images), buildings (750 images), car_rear (526 images), face (450 images), leaf (186 images) and motorbike (826 images).

B. Experimental Setup

In the following experiments, along with traditional feature extraction methods, we also employ the *bag-of-visual-features (BoVF)* model. The BoVF keypoints are obtained with the Harris-Laplace point detector and dense sampling, depending on the nature of the images. For texture images – **KTH-TIPS** database – we employ the first, and for outdoor images – **Corel-1000** and **Caltech** databases – we use the second one. Then, texture or color features are computed. Color features are extracted over the area of 10 pixels around the keypoints (see Section III). After detecting keypoints and computing their descriptors, the visual codebook is created by the k-means [14] clustering algorithm. The size K of the vocabulary ranges from 50 to 300 and the number of keypoints, or sampling, ranges from $50,000$ to $300,000$.

For classification, *Support Vector Machines (SVM)* [22], *k-means* [14] and *Nearest Neighbors (K-NN)* [23] with Euclidean distance measure are used. The *LibSVM* implementation [24] is used to train the classifier. For the K-NN classifier, a value of $K = \{10, 30, 50, 70\}$ is set. For the k-means, the number of clusters is set to the number of classes of each dataset. We employ the SVM with two kernels, linear and radial basis function (RBF), for multi-class classification. For each vocabulary size, the cost parameter is optimized using 10-fold cross validation with a parameter range of 2^{-10} through 2^{10} with 10 as a step in logarithmic scale.

C. Results

Experiment 1: First, we evaluate our approach for texture classification and compare it to the silhouette results obtained from different projections. Figure 6 shows the classification rates for different classifiers and the silhouette coefficient of the resulting projections. Features are computed with the co-occurrence descriptor for 5 distinct values of the distances parameter (1 to 5). We also investigate the performance for ALL features combined. The performance of the classifiers and silhouette coefficients are indicated in the y axis, while the distances are indicated in the x axis. Notice that the changes in the silhouette coefficient is reflected in the performance of the classifiers. As the coefficient rises or falls the classification rates also grow or drop, accordingly. This behavior confirms our hypothesis. Distance 1 provides the best classification rates, while values ranging from 2 to 5, yield lower rates. The same occurs with the silhouette coefficient, matching our hypothesis. This result indicates that projections and their silhouettes can be used as an interesting guide to defining the distance parameter of the co-occurrence feature extraction algorithm.

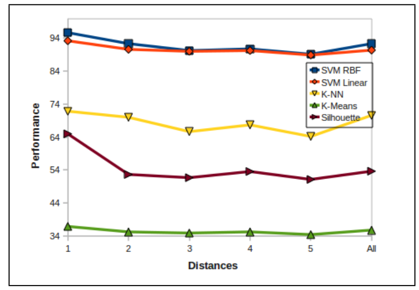
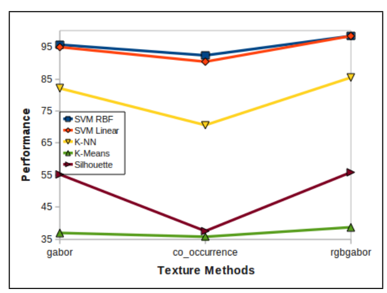
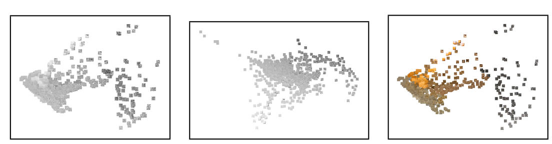


Figure 6. **KTH-TIPS** database evaluation influenced by distances of the co-occurrence matrix descriptor.

We also evaluate our approach for other traditional texture analysis methods. Figure 7 shows the results for three different descriptors: Gabor filters ($gabor$), co-occurrence matrix ($co\_occurrence$) and Gabor filters applied on the R, G, B plane of the image ($rgbgabor$). We can observe that $rgbgabor$ is superior for both classification rate and silhouette coefficient. There is a clear distinction in performance between $co\_occurrence$ (ALL distances) and $rgbgabor$. The accuracy of SVM with RBF kernel for $co\_occurrence$, $gabor$ and $rgbgabor$ is 90.37%, $94,93\%$ and 98.39%, respectively. The silhouette coefficient also reflects this accuracy, yielding values -0.3727, -0.2535 and -0.2505. The corresponding projections are shown in Figures 7(b), 7(c) and 7(d). As we move from left to right, we perceive images with coarser textures.



(a)



(b) Gabor (c) Co-occurrence (d) RGBGabor

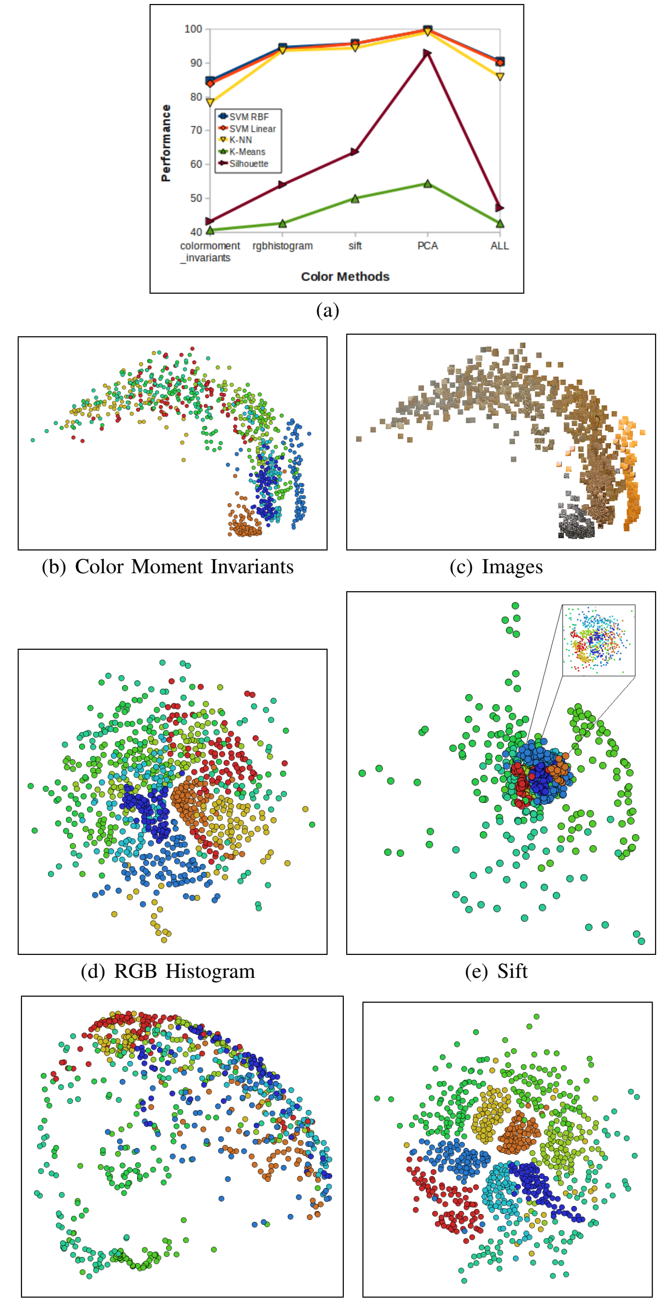
Figure 7. Comparing different sets of features obtained using traditional texture analysis methods extracted from the **KTH-TIPS** database. Again the classification rates matches the silhouette coefficients, showing the generality of our approach.

In order to investigate the contribution of the texture and color approaches for the separation of the images on the **KTH-TIPS** database, we apply the BoVF model with dense sampling and three color descriptors: $colormoment\_invariants$ [25], $rgbhistogram$ and $sift$ [26]. In Figure 8(a), the classification rates and silhouette coefficients are presented in the y axis, while the descriptors are presented in the x axis. It can be observed that $colormoment\_invariants$ do not perform well. $Rgbhistogram$ exhibits a slight improvement, followed by the $sift$ method. Their projections are illustrated in Figures 8(b), 8(d) and 8(e), respectively. Figure 8(c) shows the same projection of Figure 8(b), with circles replaced by the images they represent. To overcome the visual overlapping the center of the projection is zoomed in and shown at the top right-hand corner with smaller points in Figure 8(e).

Normally, for more complex image datasets it is desirable to combine different feature extraction approaches to capture as much as possible the underlying information present on the images. However, the resulting feature spaces can be of very high dimensionality, which in most cases impair the precision of classification tasks. We then combine the three previous descriptors and introduce an intermediate dimensionality reduction step in our approach with *Principal Component Analysis (PCA)* [27] method, reducing the space to 10 dimensions. The nearest-neighbor classification rates are $85,92\%$ and 99.12% for the original and the reduced space, matching with the silhouette cofficient, $-0,2547$ and 0.1290, respectively. Figures 8(f) and 8(g) present the projection for the combined descriptors and the intermediate step using PCA. Finally, it is worth noting that nearest neighbor classification has a higher accuracy compared to the SVM classifier. This may be associated to the good separability seen in the projections of the Figures 8(d), 8(e) and 8(g).

Experiment 2: In this experiment, we present a comparative study of our approach for evaluating local and global features. The RGB histogram descriptor is employed for both approaches. Local features are computed using the BoVF model. Figure 9 shows the classification rates in the y axis, while global and local features are indicated in the x axis. We optimize two parameters of the BoVF model: the vocabulary size and the number of keypoints. Figure 9(a) and 9(b) show the results for vocabulary sizes with $50,000$ and $100,000$ keypoints, respectively.

It can be observed that the local approach performs significantly better than the global one. The highest silhouette coefficient ($K = 250$) is 0.0055 for $50,000$ keypoints and -0.0022 for $100,000$ keypoints. The $global\_rgbhistogram$ silhouette coefficient significantly reduce to -0.1029. The same behavior is observed for the classification rates, which are 79.3% for $50,000$ keypoints and 79.78% for $100,000$, reducing to 64.30% for $global\_rgbhistogram$. With the best K at hand, we perform two others experiments: $200,000$ and $300,000$ keypoints. The result shows that our approach



(a)

(b) Color Moment Invariants

(c) Images

(d) RGB Histogram

(e) Sift

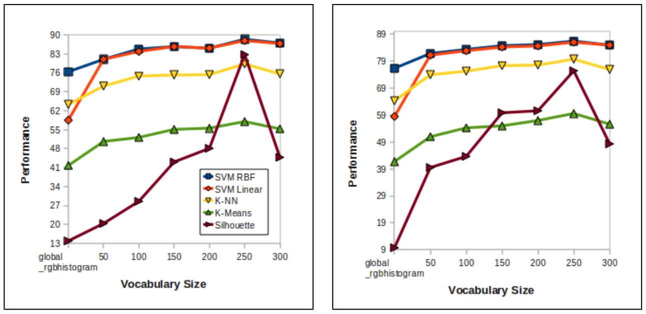
(f) Combined (ALL)

(g) Dimensionality Reduction

Figure 8. Projections of different feature spaces extracted from **KTH-TIPS** database. Applying the PCA to reduce the dimensionality results in better classification rates and silhouette values, matching with class separability observed on the produced projection (colors indicates the classes).

is consistent, being a useful method to reduce the time spent to determine the better vocabulary size on BoVF model.

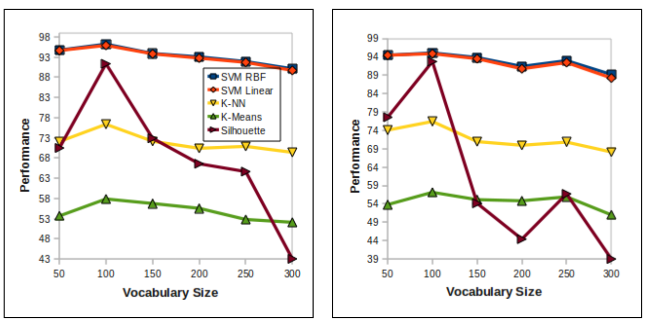
Experiment 3: In this experiment we again evaluate the outcome vocabulary size of the BoVF model regarding the classifications rates and silhouette coefficient. However, in addition to the RGB histogram descriptor, we



(a) 50,000 keypoints (b) 100,000 keypoints

Figure 9. Results for vocabulary sizes of BoVF model considering $50,000$ keypoints and $100,000$ keypoints for the **Corel-1000** database.

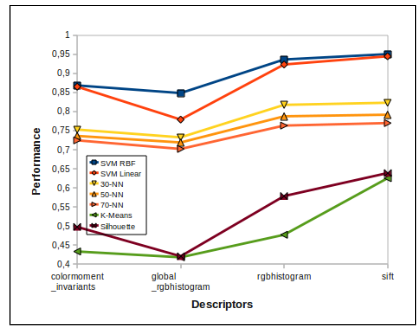
use two different methods: $colormoment\_invariants$ and $sift$. Figure 10 shows the classification rates and silhouette coefficients in the y axis, and the vocabulary size in the x axis. Figure 10(a) and 10(b) show the results for $50,000$ and $100,000$ keypoints using the $sift$ method. The highest silhouette coefficient and classification rate is obtained for $K = 100$.



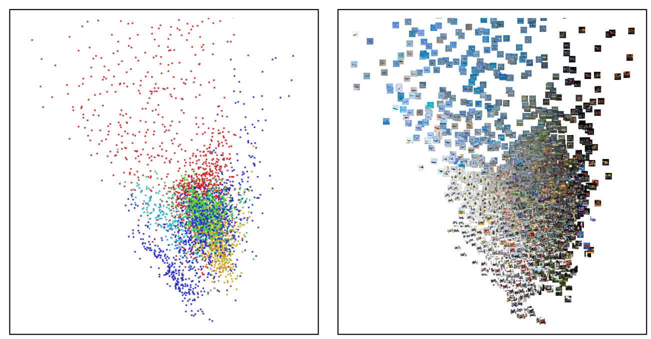
(a) $50,000$ (b) $100,000$

Figure 10. Results for different vocabulary sizes of BoVF model considering $50,000$ keypoints and $100,000$ keypoints for the **Caltech** database.

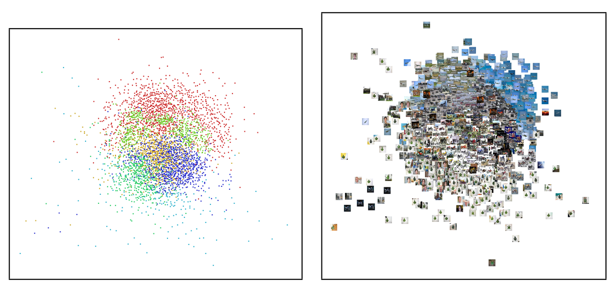
We also evaluate other color methods to build the BoVF model, taking $K = 100$ as the best value for the vocabulary size. Figure 11 shows the classification rates and silhouette coefficient for three descriptors: $colormoment\_invariants$, $rgbhistogram$ and $sift$. In addition, features are extracted by the $global\_rgbhistogram$ method. We can see that $global\_rgbhistogram$ attain the worst results. The $colormoment\_invariants$ features are slightly better. For the local $rgbhistogram$ an increase in performance is observed. All classification rates follow the silhouette coefficient obtained from the projection of the feature sets. $Sift$ achieves the best performances with $95,42\%$. For the K-NN classification, 30, 50 and 70 neighbors are used. Figures 11(b) and 11(d) show the projections for the best descriptor and the worst descriptor. The projections in which circles are replaced by images are shown in Figures 11(c) and 11(e). It is worth noting that $global\_rgbhistogram$ method is known to be very influenced by the image background, which is confirmed by the projections. On the other hand, $sift$ method, which is strongly recommended to recognize object categories in the presence of various backgrounds without segmentation, render a better projection. Therefore, projections are useful tools to support the expert to make decisions about what features sets should be used.



(a)



(b) Global RGB Histogram (c) Images



(d) Sift (e) Images

Figure 11. Comparison using projections of two different feature spaces. The best projection, in terms of class separation, matches with the classification rates, indicating that projections can be useful tools to support experts about what kind of feature should be used to represent an image dataset.

VI. Conclusions

This paper has presented a novel approach based on interactive information visualization representations to explore feature spaces. The results presented here support the hypothesis that as the quality of the employed visual representations increases, the results of correct classification rates also increases. Therefore these visual representations can be successfully used as a guide to understanding the features behavior in terms of grouping the similar images and separating the dissimilar ones. We also define an index, the silhouette coefficient, to help users assess the visual quality of the projections, overcoming problems related to the subjectiveness involved in the visual analysis.

Our work does not intend to be an approach for feature selection towards finding the best set of features for classification. Instead, we are more focused on supporting experts on the task of understanding the outcome of different feature extraction algorithm, and the effect of changing their parameters. Therefore this can be considered as a pre-step to the feature selection process, where an expert can prune the possibilities of choosing the algorithms and parameters for the feature extraction, speeding-up the whole process of finding the most appropriate space to represent a set of images.

VII. Acknowledgements

The authors acknowledge the financial support of FAPESP and CNPq.

References

[1] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Academic Press, February 2006.

[2] C. C. Aggarwal, "Towards effective and interpretable data mining by visual interaction," *ACM SIGKDD Explorations Newsletter*, vol. 3, no. 2, pp. 11–22, 2002.

[3] E. Pampalk, W. Goebl, and G. Widmer, "Visualizing changes in the structure of data for exploratory feature selection," in *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2003, pp. 157–166.

[4] F. Paulovich and R. Minghim, "HiPP: A novel hierarchical point placement strategy and its application to the exploration of document collections," *IEEE Transactions on Visualization and Computer Graphics (Proceedings of Information Visualization 2008)*, vol. 14, no. 6, pp. 1229–1236, November-December 2008.

[5] C. C. Aggarwal, "A human-computer cooperative system for effective high dimensional clustering," in *KDD '01: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2001, pp. 221–226.

[6] T. Schreck, J. Schneidewind, and D. Keim, "An image-based approach to visual feature space analysis," in *WSCG 2008, 16th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, Plzen, Bory, Czech Republic, February 2008, pp. 223–230.

[7] J. F. Rodrigues Jr., C. A. B. Castañón, A. J. M. Traina, and C. Traina Jr., "Using efficient visual exploration techniques to evaluate features for content-based image retrieval," in *XVI Brazilian Symposium on Computer Graphics and Image Processing*. Los Alamitos, CA, USA: IEEE Computer Society, 2003, pp. 183–190.

[8] J. F. Rodrigues Jr., A. J. M. Traina, and C. Traina Jr., "Enhanced visual evaluation of feature extractors for image mining," in *The 3rd ACS/IEEE International Conference on Computer Systems and Applications*, vol. 1. Washington, DC, USA: IEEE Computer Society, 2005, pp. 45–I.

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

[10] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, no. 2, pp. 1–60, April 2008.

[11] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.

[12] F. Bianconi and A. Fernández, "Evaluation of the effects of gabor filter parameters on texture classification," *Pattern Recognition*, vol. 40, no. 12, pp. 3325–3335, 2007.

[13] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1-2, pp. 43–72, November 2005.

[14] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1. University of California Press, 1967, pp. 281–297.

[15] F. Paulovich, L. Nonato, R. Minghim, and H. Levkowitz, "Least square projection: a fast high precision multidimensional projection technique and its application to document mapping," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 3, pp. 564–575, May/Jun 2008.

[16] T. F. Cox and M. A. A. Cox, *Multidimensional Scaling*, 2nd ed. Chapman & Hall/CRC, 2000.

[17] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2005.

[18] P. Brodatz, *Textures: A Photographic Album for Artists and Designers*. New York: Dover Publications, 1966.

[19] E. Hayman, B. Caputo, M. Fritz, and J.-O. Eklundh, "On the significance of real-world conditions for material classification," in *ECCV04 - 8th European Conference on Computer Vision*, vol. 4, 2004, pp. 253–266.

[20] J. Li and J. Z. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1075–1088, 2003.

[21] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2. Los Alamitos, CA, USA: IEEE Computer Society, June 2003, pp. 264–271.

[22] V. N. Vapnik, *The nature of statistical learning theory*. New York, NY, USA: Springer-Verlag, 1995.

[23] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, pp. 21–27, 1967.

[24] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001.

[25] F. Mindru, T. Tuytelaars, L. Van Gool, and T. Moons, "Moment invariants for recognition under changing viewpoint and illumination," *Computer Vision and Image Understanding*, vol. 94, no. 1-3, pp. 3–27, 2004.

[26] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[27] I. T. Jolliffe, *Principal Component Analysis (second edition)*. New York, NY, USA: Springer-Verlag, October 2002.