A Multi-class Kernel Alignment Method for Image Collection Summarization

Jorge E. Camargo and Fabio A. González

Bioingenium Research Group
National University of Colombia
{jecamargom,fagonzalezo}@unal.edu.co

Abstract. This paper proposes a method for involving domain knowledge in the construction of summaries of large collections of images. This is accomplished by using a multi-class kernel alignment strategy in order to learn a kernel function that incorporates domain knowledge (class labels). The kernel function is the basis of a clustering algorithm that generates a subset, the summary, of the image collection. The method was tested with a subset of the *Corel* image collection using a summarization quality measure based on information theory. Experimental results show that it is possible to improve the quality of the summary when domain knowledge is involved.

Keywords : Image collection summarization, information visualization, clustering, multi-class kernel alignment.

1 Introduction

Effective and efficient access to large collection of images is an important challenge for information retrieval. The main problem is how to find the right images based on their contents (content-based image retrieval, CBIR) [4]. A promising approach to this problem is based on visualization of the whole image collection using a 2D map metaphor. This strategy tries to exploit the human brain capacity for efficiently recognizing visual patterns, so that an ordered display of many images at the same time may help users to find the right information. The visualization is built in such a way that users can see different images distributed in the screen according to their visual similarity and can intuitively start to explore the image collection. In large collections of images it is not possible to show all images to the user due to the limitations of screen devices. Therefore, it is necessary to provide a mechanism that *summarizes* the image collection. This summary represents an overview of the data set and allows the user to start the navigation process. After building this summary, it must be shown to the user, this problem is called *image collection projection*. It is usually addressed using dimensionality reduction methods for obtaining a low dimensional representation of each image that can be projected in a 2D layout [7].

There are some works that have addressed the construction of image collection summaries. Some of them use clustering methods [10,9], similarity pyramids

E. Bayro-Corrochano and J.-O. Eklundh (Eds.): CIARP 2009, LNCS 5856, pp. 545–552, 2009.

methods [3], graph methods [2,6], neural networks methods [5], among others. In all the cases, the summarization problem is approached as a non-supervised learning problem. Typically, image clusters are identified in the collection and representative images from each cluster are chosen to compose the summary.

This paper proposes a combined supervised and non-supervised strategy for image collection summarization. The supervised part uses domain knowledge, in the form of a training set of labeled images, to build an image kernel function, which can be seen as an image similarity measure. The kernel function is based on individual kernel functions that measure image similarity according to different visual features. The individual kernels are optimally combined using a multi-class kernel alignment strategy. The combined kernel is then used as input for a k-medoids clustering algorithm and the generated medoids correspond to the image collection summary.

The reminder of this paper is organized as follows: In Section 2, the general summarization framework is presented and briefly discussed; in Section 3, the kernel-based approach for improving the summarization is described; Section 4, shows the experimental evaluation of the strategy. Finally, Section 5 presents the conclusions and future work.

2 Image Collection Summarization Framework

We aim to generate an overview of the image collection that faithfully represents the complete collection. To accomplish this objective, we propose the framework shown in Figure 1. The steps of the process are: selection of an image subset for training; extraction of image features; kernel alignment for involving domain knowledge; construction of a combination function based on the parameters found with kernel alignment; clustering using k-medoids for building an image collection summary; and application of a dimensionality reduction technique for generating a 2D visualization of the summary. On the other hand, features of the remaining images are extracted and the kernel function is calculated using the combination function obtained previously. When a new image arrives to the collection, it can be automatically classified in one of the clusters by calculating its similarity with the medoids using the combination function and the image is classified in the cluster of the most similar medoid.

3 Involving Domain Knowledge (Kernel Alignment)

Kernel functions have been successfully used in a wide range of problems in pattern analysis since they provide a general framework to decouple data representation and learning algorithms. A kernel function implicitly defines a new representation space for the input data in which any geometry or statistical strategy may be used to discover relationships and patterns in that new space. Intuitively, kernel functions provide a similarity relationship between objects being processed, so they are widely also used in similarity-based learning. In this work, we use kernel functions with a twofold purpose: first, to model a more

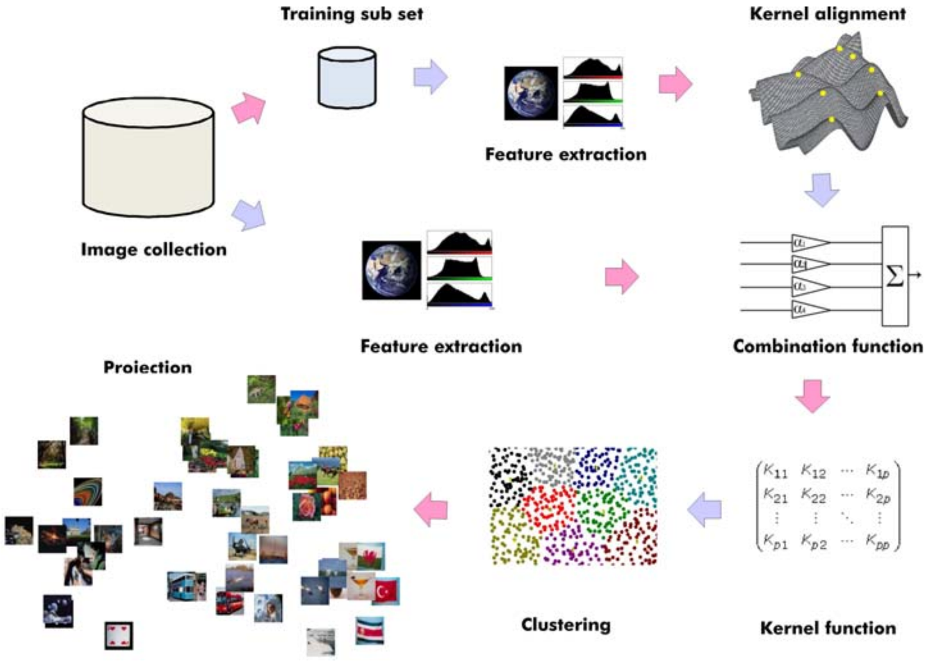

Fig. 1. Framework for summarizing an image collection

appropriate similarity measure between images using low-level visual features, and second, to learn a combination of features adapted to those particularities of the application domain.

A histogram is a discrete and non-parametric representation of a probability distribution function. Although histograms may be seen as feature vectors, they have particular properties that may be exploited by a similarity function. There are different kernel functions specially tailored to histograms. In this work, we use the *histogram intersection kernel*. Consider h as a histogram with n bins, associated to one of different visual features. The histogram intersection kernel between two histograms is defined as $k_{\cap}(h_i, h_j) = \sum_{l=1}^{n} min\,(h_i(l), h_j(l))$. This kernel calculates the area of the intersection of both histograms.

In this work, four different histograms are calculated for each image: borders, texture, RGB and gray. Using $k_{\cap}$ and these four visual features, we obtain four different kernel functions that will be used for building a new kernel. A kernel function using just one low-level feature provides a similarity notion based on particular aspect of the visual perception. For instance, the RGB histogram feature is able to indicate whether two images have similar color distributions. However, we aim to design a kernel function that provides a better notion of image similarity according to the a priori information. We construct the new kernel function using a linear combination of kernel functions associated to individual features. The most simple combination is obtained by assigning equal weights to all basis kernel functions, so the new kernel induces a representation space with all visual features. However, depending on the particular class, some features may have more or less importance.

Cristianini [8] proposed the kernel alignment strategy in the context of supervised learning to combine different visual features in an optimal way with respect to a domain knowledge target (*ideal kernel*). The empirical kernel alignment, is a similarity measure between two kernel functions, calculated over a data sample. If K_α and K_t are the kernel matrices associated to kernel functions k_α and k_t in a data sample S, the kernel alignment measure is defined as:

$$A_S(K_\alpha, K_t) = \frac{\langle K_\alpha, K_t \rangle_F}{\sqrt{\langle K_\alpha, K_\alpha \rangle_F}\sqrt{\langle K_t, K_t \rangle_F}}, \tag{1}$$

where $\langle \cdot, \cdot \rangle_F$ is the Frobenius inner product defined as $\langle A, B \rangle_F = \sum_i \sum_j A_{ij} B_{ij}$, K_α is the linear combination of basis kernels, that is, the combination of all visual features given by $k_\alpha(x, y) = \sum_f \alpha_f k_\cap \left(h_f(x), h_f(y)\right)$, where $h_f(x)$ is the f-th feature histogram of image x, and α is a weighting vector. The definition of a target kernel function K_t, i.e. an ideal kernel with explicit domain knowledge, is done using labels associated to each image that are extracted from previous information (class labels). It is given by the explicit classification of images for a particular class using y_n as the labels vector associated to the n-th class, in which $y_n(x) = 1$ if the image x is an example of the n-th class and $y_n(x) = -1$ otherwise. So, $K_t = yy'$ is the kernel matrix associated to the target for a particular class. This configuration only considers a two-class case. We need to build a new kernel function that takes into account the information of all classes simultaneously (*multi-class* case).

Vert [12] proposes a strategy that addresses the multi-class problem in the context of multi-class classification. Therefore, we adapted his strategy in the context of image collection summarization. Author proposes to build the ideal kernel matrix as follows:

$$K_t(x, x) = \begin{cases} 1 & \textit{if } \text{y=y'} \\ -1/(Q-1) & \text{otherwise} \end{cases}, \tag{2}$$

where Q is the number of classes. K_t is, by construction, a valid kernel and we will call it the *ideal kernel*. Under some regularity assumptions on K_α, the alignment function is differentiable with respect to α. Upon this assumption we can use a *gradient ascent* algorithm in order to maximize the alignment between the combined kernel and the ideal kernel as follows:

$$\alpha^* = \underset{\alpha \in \Theta}{argmax}\, A_S(K_\alpha, K_t) \tag{3}$$

Due to the fact that we have a function composed of a vector of variables, we have a *gradient vector* composed of partial derivatives $\nabla \alpha A_S = \left[\frac{\partial A_S}{\partial \alpha_1}, \frac{\partial A_S}{\partial \alpha_2}, \ldots, \frac{\partial A_S}{\partial \alpha_d}\right]^T$. The optimization algorithm starts from a random α, and at each step, updates α, in the direction of the gradient $\Delta\alpha_i = \eta \frac{\partial A_s}{\partial \alpha_i}$, $\forall_i$ and $\alpha_i = \alpha_i + \Delta\alpha_i$, where η is called the *stepsize*, or *learning factor* and determines how much to move in that direction [1].

Kernel alignment strategy has been used in the context of supervised learning and in classification problems. We use it for both, supervised and non-supervised learning in the context of summarization of collection of images.

4 Experimentation

Our main goal in this experimentation was to measure the quality of the summary. We used the Corel data set, which is a collection of photographic stock images and clip art, and it is the most widely used standard data set for testing content based image retrieval systems CBIR. A subset of 2,500 images was selected, which has 25 classes with 100 images each one (aviation, beach, cats, cards, birds, flags, forest, among others). The extracted features were Gray Histogram, RGB color histogram, Sobel Histogram (borders) and Local Binary Partitions (texture). These four visual features were modeled as discrete probability distributions and the kernel function chosen was the histogram intersection. The summary was created with a k-medoids clustering algorithm. For projecting (2D visualization), the original high-dimensional space of the image summary was projected in a low-dimensional space using Multidimensional Scaling (MDS) [11].

4.1 Summary Quality Evaluation

A good summary corresponds to representative set of samples from the collection, i.e., a set that includes prototypical images from the different categories present in the collection. Based on this idea, we define a supervised summarization quality measure that makes use of the image labels. This measure corresponds to the entropy of the summarization and is calculated as follows:

$$H_{summary} = -\sum_{i=1}^{C} (\frac{\#C_i}{k}) log_2(\frac{\#C_i}{k}), \tag{4}$$

where C is the number of classes, $M = \{m_1, \ldots, m_k\}$ is the set of k medoids obtained in the clustering process, and $\#C_i = \|\{m_j \in M | m_j \in C_i\}\|$ is the number of medoids in M that belong to class C_i. The quantity $\frac{\#C_i}{k}$ represents the proportion of samples in the summary that belongs to class C_i. The maximum entropy is obtained when this value is the same for all classes, e.i., $\forall i$, $\frac{\#C_i}{k} = \frac{1}{C}$. In this case, all the classes are equally represented in the summary. The maximum entropy depends on the number of classes, $H_{summary} = log_2(C)$. In this experimental setup $log_2(C) = log_2(25) = 4.64385619$. With this measure defined, we aim to assess the quality of the summaries generated for the following kernel functions: an *ideal kernel function* using the Equation 2, which will have the maximum entropy since it has the a priori class labels information; a *basis kernel function* as a combination of the base kernel functions (RGB, Sobel, LBP and Gray) with equal weights (alphas); and the *aligned kernel* built as was suggested in Section 3.

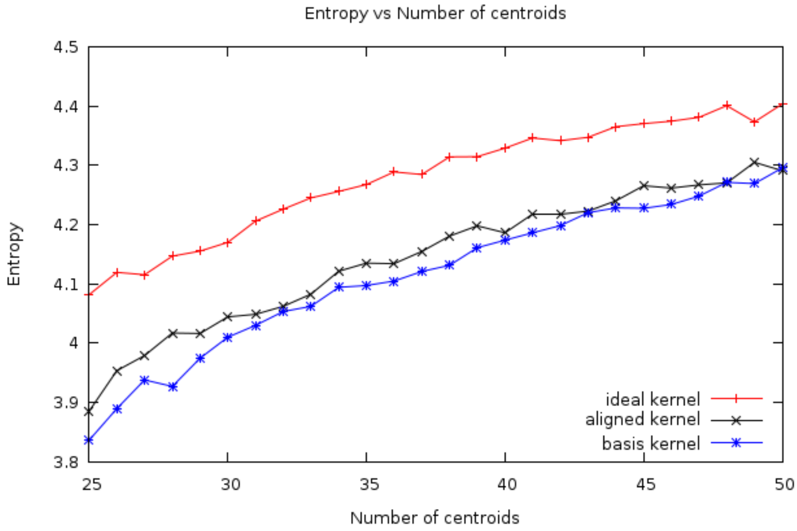

Fig. 2. Entropy vs number of centroids (average for 100 runs). The kernel that involves domain knowledge (*aligned kernel*) outperforms the base kernel.

4.2 Experimental Results

For learning the kernel function, we start the gradient ascent algorithm with α values (one per visual feature) generated randomly (50 times), $\eta = 0.1$, $\nabla\alpha_i = 0.1$ and 100 iterations. Table 1 shows the α values obtained using gradient ascent for optimizing the kernel alignment, which indicates that color feature (RGB) has the highest weight in the combination function. It is because images of the Corel data set have similar color distribution in each class (in other data sets it would be different). On the other hand, texture feature (LBP) has the lowest weight, which indicates that texture is not a good class discriminant in this data set. Figure 2 shows the quality of the three summaries: *ideal kernel, basis kernel,* and *aligned kernel.* Results show that the *aligned kerne*l outperforms the baseline, which proves that the proposed method improves the quality of the summary. All three kernel increase the summary entropy when the number of medoids is increased; with k=*50* medoids the summary entropy is close to the maximum. Figure 3 shows the visualization (2D projection) of the entire Corel data set using MDS with the medoids of each cluster. Figure 4 shows a visualization of the the summary, which involves

Table 1. Alpha values found for the combination function obtained with multi-class kernel alignment

Feature	GRAY	LBP	SOBEL	RGB
α	0.1537	0.0507	0.1023	0.6932

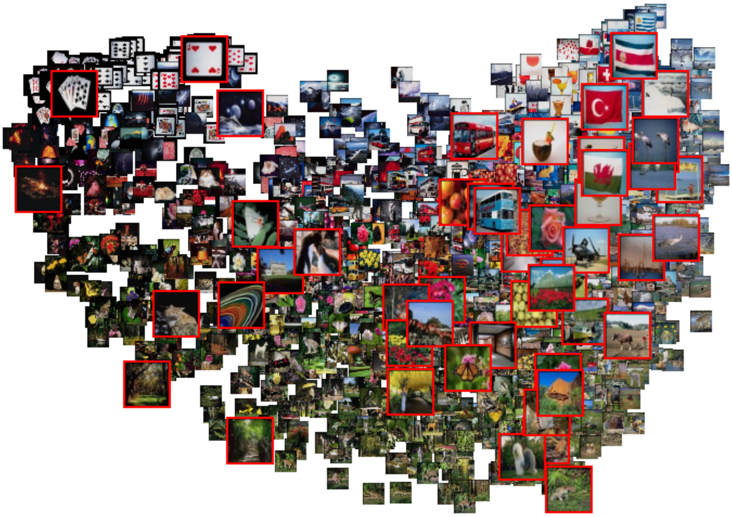

Fig. 3. Visualization of the Corel collection with 50 medoids highlighted

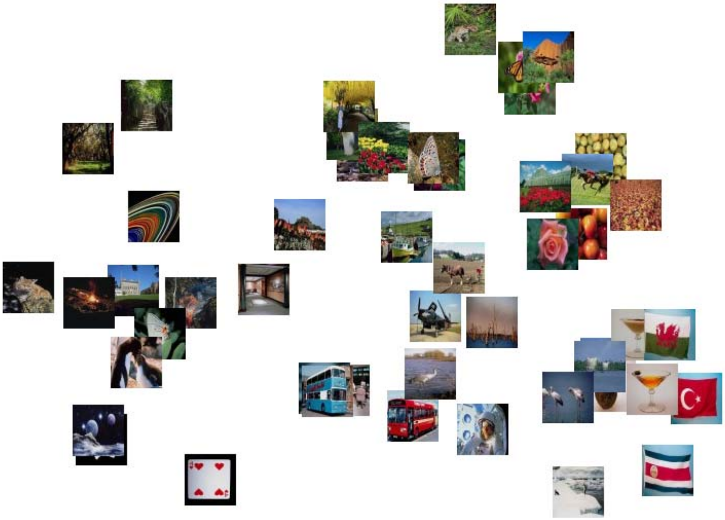

Fig. 4. Visualization of the image collection summary

domain knowledge and represents the entire collection with a higher precision than a summary without domain knowledge.

5 Conclusions and Future Work

We have presented a method for involving domain knowledge in the construction of image collection summaries. We use kernel alignment strategy for both, supervised and non-supervised learning in the context of summarization of large collections of images. We propose a quantitative measure based on information theory for assessing the quality of the summary. Results show that the summary is improved when it is built following the proposed method. With the model proposed in this work, it is possible to automatically classify a new image that arrives to the collection in one of the clusters by calculating the combination function for the new image and calculating its similarity with respect to the images of the summary. In future work, we will evaluate other clustering techniques and we will assess the strategy with real users using quantitative and qualitative measures that allow us to fit the model.

References

1. Alpaydin, E.: Introduction to Machine Learning. MIT Press, Cambridge (2004)
2. Cai, D., He, X., Li, Z., Ma, W.-Y., Wen, J.-R.: Hierarchical clustering of www image search results using visual, textual and link information. In: Proceedings of the 12th annual ACM international conference on Multimedia, pp. 952–959 (2004)
3. Chen, J.-Y., Bouman, C.A., Dalton, J.C.: Hierarchical browsing and search of large image databases. IEEE Transactions on Image Processing 9(3), 442–455 (2000)
4. Joshi, D., Li, J., Wang, J.Z., Datta, R.: Image retrieval: Ideas, influences, and trends of the new age. ACM Comput. Surv. 40(2), 1–60 (2008)
5. Deng, D.: Content-based image collection summarization and comparison using self-organizing maps. Pattern Recognition 40(2), 718–727 (2007)
6. Gao, B., Liu, T.-Y., Qin, T., Zheng, X., Cheng, Q.-S., Ma, W.-Y.: Web image clustering by consistent utilization of visual features and surrounding texts. In: MULTIMEDIA 2005: Proceedings of the 13th annual ACM international conference on Multimedia, pp. 112–121. ACM, New York (2005)
7. Nguyen, G.P., Worring, M.: Interactive access to large image collections using similarity-based visualization. Journal of Visual Languages & Computing 19(2), 203–224 (2008)
8. Shawe Taylor, J., Cristianini, N.: Kernel Methods for Pattern Analysis. Cambridge University Press, Cambridge (2004)
9. Simon, I., Snavely, N., Seitz, S.M.: Scene summarization for online image collections. In: IEEE 11th International Conference on Computer Vision, 2007 (ICCV 2007), pp. 1–8 (2007)
10. Stan, D., Sethi, I.K.: eid: a system for exploration of image databases. Inf. Process. Manage. 39(3), 335–361 (2003)
11. Torgerson, M.S.: Multidimensional scaling: I. theory and method. Psychometrika 17(4), 401–419 (1958)
12. Vert, R.: Designing a m-svm kernel for protein secondary structure prediction. Master's thesis, DEA informatique de Lorraine (2002)