Context-based Image Similarity Queries*

Ilaria Bartolini

DEIS - IEIIT-BO/CNR, University of Bologna, Italy
ibartolini@deis.unibo.it

Abstract. In this paper an effective context-based approach for interactive similarity queries is presented. By exploiting the notion of image "context", it is possible to associate different meanings to the same query image. This is indeed necessary to model complex query concepts that, due to their nature, cannot be effectively represented without contextualize the target image. The context model is simple yet effective and consists of a set of significant images (possibly not relevant to the query) that describe the semantic meaning the user is interested in. When feedback is present, the query context assumes a dynamic nature, changing over time depending on the actual retrieved images judged as relevant by the user for her current search task. Moreover, the proposed approach is able to complement the role of relevance feedback by persistently maintaining the query parameters determined through user interaction over time and ensuring search efficiency. Experimental results on a database of about 10,000 images show the high quality contribution of the proposed approach.

1 Introduction

Advances in the computer technologies and the advent of the Word-Wide Web have produced the explosion of an increasing number of complex data such as digital images, video, and audio. As a primary consequence, there is a pressing need for the definition of efficient and effective techniques able to retrieve such information based on their content.

The traditional paradigm for the retrieving of images is based on keyword annotation. In this approach, human experts manually annotate each image with a textual description, so that text-based information retrieval techniques can be applied [6]. This approach has the advantage of inheriting efficient technologies developed for text retrieval, but is clearly impracticable for the case of very large image DBs. Moreover, its effectiveness highly depends on the subjective opinions of the annotators, who are also likely to supply different descriptions for the same image.

To overcome above difficulties, in the early 1990's an alternative approach has been proposed. Content-Base Image Retrieval (CBIR) uses visual properties (*features*) to represent the image content. This approach has a wider applicability, since features can be computed automatically, and the information used during the retrieval process is always consistent, since it does not depend on human interpretation. To characterize each image, CBIR systems define a set of low level relevant features able to effectively characterize the content of the images and then use such features for retrieval

* This work is partially supported by the WISDOM MIUR Project

purposes [8]. The features should be "simple enough" to allow the design of automatic extraction algorithms, yet "meaningful enough" to capture the image content. Under this view, each image is typically represented by a high-dimensional *feature vector*, whose dimensionality depends on the number and on the type of extracted features, and similarity between images is assessed by defining a suitable distance function on the resulting feature space.

CBIR systems, however, assume that high level concepts (as perceived by humans) can be perfectly mapped to low level features (as extracted by the computer): This, of course, may be not always true. This mismatch between human-perceived and computer-provided image representations is known as the "semantic gap" problem [8] and is one of the most challenging problem for multimedia information retrieval.

Although approaches based on a-priori classification of images [9] and on analysis of (possibly available) surrounding text/captions [10, 14] might help in alleviating the semantic gap, they are not always applicable for heterogenous image collections. Moreover, such approaches are not able to "contextualize" the search based on current user needs. Indeed, even a same image might represent different meanings to different users (or to a same user at different times). For example (see [7]), a portrait can suggest the notion of "painting", when placed in the context of other painting images, and the meaning of "face", when the context becomes a set of people photos.

Motivated by above observations, in this paper we investigate the potentialities of an approach to contextualize image queries with the aim to solve, or at least alleviate, the semantic gap problem. The key idea is to complement the image query with a set of (possibly not relevant) images able to direct the search to the correct semantic concept (see Section 3 for a real example). Even if our approach is very simple, it is indeed effective and does not require neither a-priori classification, nor textual information associated to images. Furthermore, it can easily complement available feedback techniques [3] by providing a better starting point for the search of complex semantic concepts and a beneficial inertial behavior on the updating of query parameters in the user-system interaction process. This is made possible by exploiting the history of the "best" relevant examples over time. Finally, to ensure search efficiency, our solution can be easily integrated with techniques for learning user preferences (e.g., [1, 13, 15]), by maintaining optimal parameters for each user query. In this way, the main limit of traditional relevance feedback techniques, consisting in "forgetting" user preferences across multiple query sessions, is no more a concern.

Our experiments, conducted on a dataset of about 10,000 images, show that the quality of results obtained from our approach, as measured in term of classical *precision*, outperforms modern interactive retrieval techniques. As for the efficiency, we integrate our context-based method in FeedbackBypass [1], and experimentally prove how the number of search interactions needed to reach a given level of precision is reduced.

The rest of the paper is organized as follows. Section 2 surveys some approaches to context-based image similarity. In Section 3 we describe our approach by defining the notion of context. The case when user feedback is present is contemplated in Sections 4 and 5, where an accurate description of the context updating is also provided. In Section 6 we present experimental results showing the effectiveness and the efficiency of

our approach. Finally, Section 7 concludes the paper and suggests directions for future work.

2 Context-based Queries

To the best of our knowledge, no other work has attempted to analyze the effect of using the notion of context, as a set of possibly not relevant images evolving over time, for content-based similarity queries. However, many works share our main goal (i.e., to alleviate, if not completely solve, the semantic gap problem) even if they usually associate a different meaning to the word "context". In this section, thus, we only survey the contributions of such works, classifying them into three main classes:

Analysis of surrounding text: Here the context is defined as the description of the image content that comes from sources other than its visual properties. Typically such content is expressed in term of *textual* information (e.g., [10, 14]) that comes from manually annotations (e.g., keywords, descriptions, etc.), or surrounding text that is available with the image (e.g., captions, nearby text from Web pages containing the image, subtitles, etc.). The similarity between images is then assessed by also taking into account similarity between associated texts, using standard text retrieval techniques [6]. In details, in [14] the authors propose an image retrieval system that combines visual (i.e., content) and textual (i.e., context) querying at a semantic level, finding a semantic association between low level features and high level concepts. Authors in [10] do the same by also integrating in the search process the notion of *form* of a multimedia document defined as the internal structure of a document (e.g., objects of an image, frames in a video, and chapters of a book).
However, in general a textual information might not be available for every image, or it might not be meaningful to correctly describe the image. This represents the main limit of this context definition.

Taxonomy-based search: The context is represented by means of subject classes (i.e., an ontological concept that represents the semantic content of an image) and by the corresponding definition of a *taxonomy* that arranges such classes into a is-a hierarchy. In this scenario, the search process starts by first browsing the taxonomy, then a classical content-based retrieval is applied. In particular, the WebSeek system [9] provides a powerful semi-automatic approach for classifying images and videos on the Web according to a general subject taxonomy based on text associated with those images and videos, as, for example, Web addresses (or URLs), HTML tags, and hyperlinks between them. Thus, a user looking for images of comets has to browse the "astronomy" category in the taxonomy, and eventually reach the "astronomy/comets" class.
This approach suffers the same limitations above described, in that a classification may not always be available. Furthermore, due to different meanings that a same image/video can assume depending on the particular user and context in which it is collocated, it would be natural to assign it to different semantic classes. The main consequence is that if the selected class in the first phase is not correct, the result of the final search is not accurate.

Analysis of objects within a same image: This is the notion of context usually assumed in research areas of image retrieval, such as computer vision. In this case, images

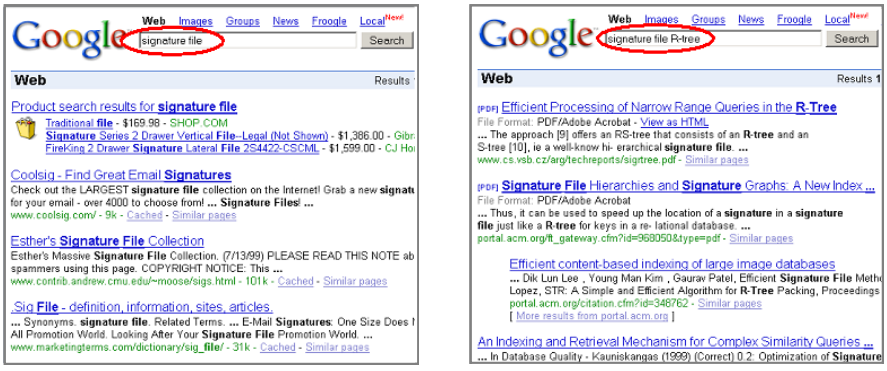
are characterized by means of local pictorial objects and their spatial relationships. This is due to the fact that such relationships are able to capture the most relevant part of semantic information in an image. Relation-based representations require that objects are represented by symbols [11]. Each object is replaced by a symbolic label located at a set of object representative points. In this way, both spatial and topological relations are supported. Relevant examples are: Object A is on object B, A is on the right/left side of B, A disjoints B, A contains B, etc.. This usually requires the use of object recognition techniques that severely limit the applicability of this approach to the case of large heterogeneous image collections, i.e., the only ones we take into consideration.

The El Niño image browsing interface [7], even if it does not represent an image retrieval system, is also relevant to our work. When browsing with El Niño, the user implicitly defines semantic concepts by selecting relevant images from a set of randomly displayed objects, and by placing them on the screen so as to reflect their (intended) mutual similarities. This is similar to the definition of our context, since the system adapts its (internal) similarity criterion by using the placement of images, so that the updated display of images matches the similarity intended by the user.

3 Our Approach

The basic idea of the proposed approach is to contextualize user queries by associating a set of objects (defining a *context*) that are possibly not relevant to the queries but that can be helpful to the system in solving, or at least alleviating, the semantic gap problem.

To give an intuitive example, let us consider a typical Web search scenario where a user is looking for documents related to “access methods” of type “signature file”. If the user adopts the well known search engine *Google*[1] and enters the keywords “signature file”, she obtains the results shown in Figure 1 (a). As it can be observed, none of the



(a) (b)

Fig. 1. Google results for the general concept “signature file” (a) and the contextualized concept “signature file R-tree” (b).

[1] Google: `http://www.google.com`.

retrieved documents is relevant with respect to the query.[2] This is due to the different meanings associated to the string "signature file", i.e., a file containing a signature or an access method. Thus, the problem is: How to tell the system that we are looking for access methods and not for file of signatures? A simple yet effective solution is to add some additional terms, e.g., "R-tree", that are not directly relevant to the query, but that helps to define a context, i.e., access methods (see Figure 1 (b) for the contextualized result).

In the image domain, the idea is to start from an image query complemented by some other (possibly not relevant) images that contextualize the query by associating the right semantic concept. Even if the approach is simple, it is indeed an effective query model, more flexible with respect to the usual Query By Example (QBE) paradigm. In this scenario, in fact, the same query can be used to retrieve different semantic concept images. This is particular important when user preferences are exploited during the search (i.e., when applying relevance feedback techniques). A simple example is shown in Figures 2, where the same "sheep" image query is used to formulate two completely different searches. This is possible by defining two contexts (in the example, these are represented by means of three images): $Context_1$ (Figures 2 (a)) represents the concept "forest", whereas $Context_2$ (Figures 2 (b)) the concept "domestic images". It is possible to observe how, in this particular examples, none of the context images are relevant to the query concepts (i.e., none of them represent animals). However, context images allow a better definition of the "query parameters" used in the search process (as confirmed by experimental results in Section 6).

(a) (b)

Fig. 2. The same query associated to a context that suggests (a) "animals in the forest" and (b) "domestic animals".

More precisely, we frame our discussion in the context of the *vector space* model, where an image is represented by a point $\mathbf{p}$ in a D-dimensional space of features, $\mathbf{p} = (p[1], \ldots, p[D])$. Given two points, $\mathbf{p}$ and $\mathbf{q}$, their similarity is measured by means of a distance function d on such space. In details, we adopt the weighted Euclidean distance ($L_{2\mathbf{W}}$, with $w[i] \in \mathbf{W}$ defaults to 1, $\forall i \in [1, D]$).

The usual approach for a user is to submit a query $Q = (\mathbf{q}, k)$, where $\mathbf{q}$ is the query point and k represents the number of results to be returned by the system. Using

[2] For lack of space, we report in Figure 1 the first four documents of the complete result; however, the same behavior is also confirmed by the other, not shown, documents.

a default distance function d, $\mathbf{q}$ is compared with the database objects and the k objects which are closest to $\mathbf{q}$ according to d are returned (i.e., $Result(Q, d) = \{\mathbf{p}_1, \ldots, \mathbf{p}_k\}$). Thus, the query point $\mathbf{q}$ and the default weights determine the final results.

However, as also shown in the examples of Figure 2, when the query concept becomes more complex, a single object (i.e., the query point $\mathbf{q}$) is not able to represent it well. To overcome such limitation, we propose a new query model where the notion of *context* plays the main role. In the new scenario, a query Q is defined as a triple $Q = (\mathbf{q}, \mathcal{C}, k)$, where $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_m\}$ is a set of points in the feature space (possibly not relevant to the query) that contextualize $\mathbf{q}$. The problem faced by the context-based approach can be precisely formulate as follows:

Problem 1 *Given a query point $\mathbf{q}$ and a context $\mathcal{C}$, determine the distance function $d^{\mathcal{C}}(\equiv \mathbf{W}^{\mathcal{C}})$ able to reflect the intrinsic meaning of images in $\mathcal{C}$. The equivalence highlights that $d^{\mathcal{C}}$ is the distance function obtained when the weights are set to $\mathbf{W}^{\mathcal{C}}$.*

The problem, thus, consists in reshaping the distance function in order to obtain more accurate results wrt those retrieved using the default distance d. The rationale is that objects in $\mathcal{C}$ are possibly not relevant; thus, they are excluded in the definition of the query point, that remains unaltered, but they are involved in the distance function re-weighting process.

4 Context Evolution

In the previous section we have focused on the first round of the search process. We now consider the possibility, for the user, to provide an evaluation of the relevance of the result objects. In particular, we suppose that, after receiving the result for her query, the user provides her positive/negative feedback on the objects in $Result(Q, d)$. Then, a new query $Q_{new} = (\mathbf{q_{new}}, k)$ and a new distance function $d_{new}(\equiv \mathbf{W_{new}})$ are computed to determine the second round of results. In this way, a feedback loop process takes place until the user is satisfied with the final result. In such context, the results of each round of search only depend on $\mathbf{q_{new}}$ and $\mathbf{W_{new}}$, that we refer to as *query parameters*.

In this scenario, we rely upon two basic strategies for learning from relevance feedback: the *query point movement* (QPM) and the *re-weighting* techniques. The former concerns the computation of an ideal query point by moving towards the positive results and away from the negative examples. A well-known implementation of this idea has been proposed by Rocchio [6] in the context of information retrieval. Given two sets $\mathcal{R}$ and $\mathcal{N}$ of relevant and not relevant objects, the new query point is adapted as:

$$\mathbf{q_{new}} = \alpha\mathbf{q} + \beta\left(\frac{1}{|\mathcal{R}|}\sum_{\mathbf{p}\in\mathcal{R}}\mathbf{p}\right) - \gamma\left(\frac{1}{|\mathcal{N}|}\sum_{\mathbf{r}\in\mathcal{N}}\mathbf{r}\right) \quad (1)$$

where α, β and γ are weighting factors, satisfying the constraint $\alpha + \beta + \gamma = 1$. More recently, QPM has been applied in several image retrieval systems [5, 3].

The re-weighting strategy updates the distance function by enhancing the feature components, which are more important than others in determining whether a result point

is relevant or not. In an early version of the MARS system [5], the authors propose a re-weighting approach based on the standard deviation of the positive examples (i.e., $w[i] = 1/\sigma[i]$). Later on, it was proven in [3] that the optimal choice of weights is to have $w[i] \approx 1/\sigma[i]^2$. The intuition behind using standard deviation and variance is that a large variance among the values of positive objects in a dimension means that the dimension poorly capture the user information need; viceversa, a small variation indicates that the dimension is important to the user expectation and should carry a higher weight.

When user feedback is present, it is reasonable to consider that the query context $\mathcal{C}$ might be refined/changed depending on the actual retrieved and relevant images found so far. The rationale is that the set of objects defining $\mathcal{C}$ have the primary role to promote an effective discover of a first set of results, i.e., to increase the number of relevant objects in the first round of search. This is indeed an important issue as it represents a better starting point for the application of relevance feedback techniques wrt the results obtained by means of a default distance. However, in our approach the context has a further important role that consists in maintaining the "inertia" of the weights during the re-weighing process, inspired by the inertial behavior of the "Rocchio-like" modification of the query point. Thus, the context has a dynamic nature and has to be properly updated over time. The main idea is to update $\mathcal{C}$ at each round of search (starting from the first round of results) by promoting the exclusion from $\mathcal{C}$ of l selected objects and the inclusion in $\mathcal{C}$ of l specific relevant objects, as dictated by policies that we will describe in the following.

With this in mind and by supposing to use only positive feedback, Problem 1 can be re-formulated as follows:

Problem 2 *Given a query point* $\mathbf{q}$*, a current context* $\mathcal{C}$ *and a current set of relevant objects* $\mathcal{R}$*, determine the optimal query parameters* $(\mathbf{q_{new}}, \mathbf{W}^{\mathcal{C},\mathcal{R}}_{\mathbf{new}})$ *for* $\mathbf{q}$*, where* $\mathbf{q_{new}}$ *is the new query point based on* $\mathcal{R}$*, and* $\mathbf{W}^{\mathcal{C},\mathcal{R}}_{\mathbf{new}}$ *are the new weights computed by means of objects in* $\mathcal{C}$ *and* $\mathcal{R}$*.*

As for the computation of the new query point, we follow the usual QPM approach (see Equation 1). Before entering into details of the re-weighting process, we first describe how the context evolves over time.

4.1 Context Switch

The context switch technique allows to update $\mathcal{C}$ depending on the relevant images in the search result. To this end, distances between context images and relevant images and between all the relevant examples are computed. Depending on the particular "selecting" policy, it is possible to assert which are the images that have to leave the context and the relevant examples that have to replace them. In particular, we provide two different policies:[3]

[3] The two strategies are inspired by the heuristic techniques of *near* and *distant expansion* proposed in [4] aiming to change the set of query points in the context of the multi-point query expansion relevance feedback approach.

Near-selection The l context images that are farthest (i.e., whose average distance is higher) to all the positive examples have to leave the context and are replaced by the l relevant examples which are closer to all the other relevant examples. The rationale is simple yet effective: Images kept in $\mathcal{C}$ over time are those which are "close" to the positive examples and, thus, have a high probability to be relevant for the search. This strategy implies that after a certain number of rounds (that is proportional to l and the cardinality of $\mathcal{C}$) the context contains only relevant images that represent the main semantic concept at a high level of granularity.

Distant-selection Conversely, the l context images that are closest (i.e., whose average distance is smaller) to all the positive examples have to leave the context and are replaced by the l relevant examples which are farthest to all the other relevant examples. The rationale here is that context images that are close to positive examples probably contain similar information; thus, they are considered redundant and are discarded. On the other hand, images that are distant to the positive examples are kept over time in $\mathcal{C}$, because they are considered more discriminant. As a consequence, there is no guarantee for images in $\mathcal{C}$ to become all relevant with respect to the semantic concept.

4.2 Context Re-weighting

Due to the dynamic nature of the context, and in order to better represent the information represented by each image in $\mathcal{C}$, we introduce a *local* context weight $w_{\mathbf{c}_j}$ for each image $\mathbf{c}_j$ that defaults to $1/m$ (m is the number of context images) and that is updated at each search round. In particular, we define two strategies for its re-computation:[4]

Maximum distance The lowest weight is associated to the image $\mathbf{c}_j$ that is more distant to the positive examples. In details, for each $\mathbf{c}_j$ the maximum distance among the positive results is computed as follows:

$$\Delta\mathbf{c}_j = \max_{\mathbf{p}\in\mathcal{R}}\{d(\mathbf{c}_j, \mathbf{p})\} \tag{2}$$

Minimum distance Conversely, the highest weight is associated to the image $\mathbf{c}_j$ that is closest to the positive examples. For each $\mathbf{c}_j$ the minimum distance among the positive results is computed as follows:

$$\Delta\mathbf{c}_j = \min_{\mathbf{p}\in\mathcal{R}}\{d(\mathbf{c}_j, \mathbf{p})\} \tag{3}$$

Then, the weight $w_{\mathbf{c}_j}$ is computed as:

$$w_{\mathbf{c}_j} = \frac{\sum_{h\neq j}\Delta\mathbf{c}_h}{\sum_h \Delta\mathbf{c}_h} \tag{4}$$

In this way, if $\Delta\mathbf{c}_j^*$ is the maximum/minimun value of $\Delta\mathbf{c}_j$, we are guaranteed that its corresponding weight $w_{\mathbf{c}_j^*}$ is the lowest/highest, respectively.

[4] The two solutions take inspiration from the *Maximum distance strategy* described in [4] aiming to solve the query re-weighting problem for multiple feature representation feedback.

Moreover, a *global* context weight $w^{\mathcal{C}}$ is computed by considering the whole set of context images. In details, the i-th weight component is:

$$w^{\mathcal{C}}[i] = \frac{|\mathcal{C}|}{\sum_{\mathbf{c}_j \in \mathcal{C}} w_{\mathbf{c}_j} (c_j[i] - \overline{c}[i])^2} \tag{5}$$

where $\overline{c} = \sum_{\mathbf{c}_j \in \mathcal{C}} c_j / |\mathcal{C}|$ is the average vector of context images.

4.3 Global Re-weighting

We are now ready to precisely describe the global re-weighting process that takes into account the contribution of both the relevant examples and the context images. In particular, we compute the global weight of the i-th coordinate as:

$$w_{new}[i] = \beta w^{\mathcal{R}}[i] + \delta w^{\mathcal{C}}[i] \tag{6}$$

where $w^{\mathcal{R}}[i]$ is the weight component computed on the current relevant examples following the re-weighting strategy proposed in [3], and $w^{\mathcal{C}}[i]$ represents the weight component derived from the context images (computed as described in the previous section). Finally, β and δ are weighting factors, satisfying the constraint $\beta + \delta = 1$.

As already mentioned at the beginning of Section 4, our re-weighting approach ensures an inertial behavior on the weights by means of the $w^{\mathcal{C}}[i]$ term. The evolution over time of $\mathcal{C}$, in fact, guarantees that relevant images that represent common (for the "near selection" strategy) or distinguishing (for the "distant selection" approach) characteristics of the relevant examples are taken into account for the re-computation of the weights in further search rounds.

5 Exploiting Prior Preferences

In the previous section, we have shown how the context-based approach to image similarity queries is able to complement available relevance feedback techniques, by providing a better starting point for the search of complex semantic concepts and by allowing an inertial behavior on the re-weighting process able to maintain the history of the "best" relevant examples. However, our approach share with all the relevance feedback methods the limit to "forget" user preferences across multiple query sessions. This means that the feedback loop has to be restarted for every new query. To overcome such limitation, many leaning user preferences techniques have been recently proposed (e.g., [1, 13, 15]) for implementing interacting similarity queries.

With particular attention to the FeedbackBypass[5] technique presented in [1], the main idea is to store and maintain the information on the query parameters gathered from past feedback loops, in order to exploit it during new query sessions, either to bypass the feedback loop completely for already-seen queries, or to start the search from a near-optimal configuration for similar images. In other words, FeedbackBypass tries

[5] FeedbackBypass: `http://www-db.deis.unibo.it/FeedbackBypass/`.

to "predict" what the user is looking for on the basis of the image query submitted to the system only. We synthetically represent this general approach as a mapping:

$$\mathbf{q} \mapsto (\mathbf{q_{opt}}, \mathbf{W_{opt}}) \tag{7}$$

which assigns to the initial image query $\mathbf{q}$ an optimal query point $\mathbf{q_{opt}}$ and an optimal set of weights $\mathbf{W_{opt}}$.

However, since FeedbackBypass only stores a single set of query parameters for each image query, the user cannot change her preferences to express a different semantics for a same query image. As also argued in [13, 15], this represents a limit because the assumption that the behavior of all users (or of the same user at different time) is the same for a given query is not always realistic (see also Figure 2).

We solve the above problem by means of our context-based approach. In details, we integrate the new query model $Q = (\mathbf{q}, \mathcal{C}, k)$ in the FeedbackBypass kernel, obtaining a new mapping definition:

$$(\mathbf{q}, \mathcal{C}) \mapsto (\mathbf{q_{opt}}, \mathbf{W_{opt}}) \tag{8}$$

which assigns to the couple $(\mathbf{q}, \mathcal{C})$ an optimal query point $\mathbf{q_{opt}}$ and an optimal set of weights $\mathbf{W_{opt}}$. Thus, associating distinct contexts to the same image, it is now possible for FeedbackBypass to support different searches that use the same query.

In our current implementation of FeedbackBypass we use a *Support Vector Machine* (SVM) for regression [12, 2] to establish the mapping between $(\mathbf{q}, \mathcal{C})$ and the relative parameters $(\mathbf{q_{opt}}, \mathbf{W}^{\mathcal{C},\mathcal{R}}_{\mathbf{opt}})$. In fact, we experimentally found that the implementation of FeedbackBypass that uses SVM is more effective than the previous one based on Wavelets.

Given a set of training points, SVM for regression estimates the shape of the unknown function by selecting, among a set of a priori known functions, the one that minimizes the average error computed in approximating all the training points. As for the choice of the kernel for the SVM, we adopt the commonly used *Radial Basis Function* (RBF).

6 Experimental Evaluation

In this section we experimentally quantify the improvement introduced by our context-based approach in the similarity query process in term of effectiveness and efficiency. Results were obtained on a dataset of about 10,000 images extracted from the IMSI collection.[6] Each image is represented in the HSV color space as a 32-D histogram, obtained by quantizing the Hue and Saturation components in 8 and 4 equally-spaced intervals, respectively.

The query workload consists of 10 randomly chosen images. For each query image, a set of volunteers defined a corresponding semantic concept (e.g., "animals in the forest", "domestic animals", "birds on the sea", etc.) and images in the dataset were classified according to such "ground truth": this allows the objective definition of relevance of an image wrt a query. Then, for each query, the users selected a context, i.e., a

[6] IMSI MasterPhotos 50,000: `http://www.imsisoft.com`.

set of images able to contextualize the query with respect to its corresponding concept. In our current implementation, the context is defined by three images and at each step the context switch involves one image change.

To measure the effectiveness of our solution we consider the classical *precision* metric, i.e., the percentage of relevant images found by a query (in our experiments we set $k = 50$), averaged over the 10 queries. As for the efficiency, we compute the precision gain introduced by FeedbackBypass with respect to the results obtained without exploiting prior user preferences (at the same level of work, i.e., number of rounds). This is to show that FeedbackBypass indeed allows to reduce the number of search rounds needed to reach a given level of precision.

The results we report refer to two main competitors:

1. Default: This is the approach currently used by all interactive retrieval systems, where the search is started by using the user query point and the default distance.
2. Context: Represents our context-based approach, where we search using a query point, a context and a distance function reflecting the meaning of the context images. In particular, depending on the strategy applied for the context switch and the context re-weighting, we specialize Context as follows:
   (a) Context-near: applies the *near-selection* and the *maximum distance* policies;
   (b) Context-distant: *distant-selection* and the *minimum distance* strategies are applied.

6.1 Experimental Results

Experiment 1: The aim of our first experiment is to measure the contribution of the context at the first round of search (i.e., without taking into account user preferences). To this end, we compare precision values obtained for the Context and the Default strategies, by considering both scenarios where FeedbackBypass is switched "off" (referred as No FB) and "on" (i.e., FB), respectively.

Results in Figure 3 confirm that Context consistently outperforms Default for the first search round. In particular, Context produces an average improvement over Default of 119.03% for the case No FB and of 24.13% for the case FB, respectively. The lower contribution for the latter case is due to the fact that here prior relevance judgements are exploited.

By analyzing the contribution of the FeedbackBypass technique in terms of search efficiency, we observe how the prior knowledge on user preferences is able to produce an improvement of precision of 176.19% for Default and of 56.52% for Context, respectively, with respect to the No FB scenario.

Experiment 2: In this second experiment our objective is to evaluate the precision dynamic (over 8 rounds) of Context-near, Context-distant and Default strategies (for simplicity of explanation we only consider the case No FB). In this scenario, we make the assumption that user feedback is present. Results shown in Figure 4 report precision vs number of rounds of search and demonstrate that both Context-near and Context-distant strategies improve over Default of 312.5% and 87.5%, respectively, at the 3th round of search and of 81.48% and 18.51%, respectively, at the 8th step. It comes out that Context-near is the winning strategy, thus in the following experiments we will

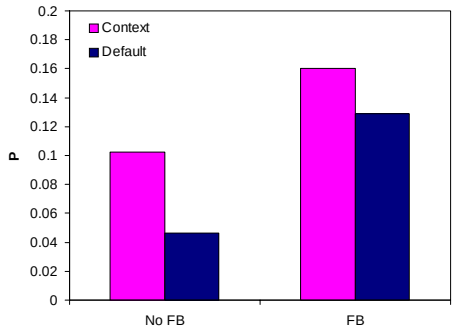


Fig. 3. Precision at the first search round for Context and Default strategies, without (No FB) and with (FB) FeedbackBypass.

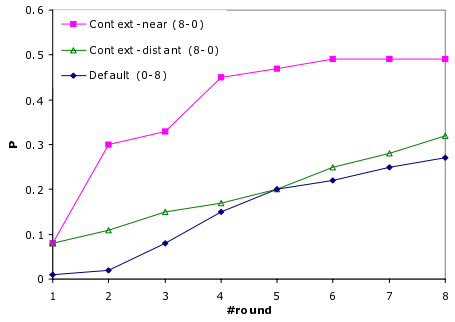


Fig. 4. Precision dynamics (8 rounds): Context-near and Context-distant strategies vs Default. The number pairs represent the number of rounds computed with and without context, respectively (e.g., (8-0): all 8 rounds executed using context).

only focus on it. Its better effectiveness is due to the adopted near-selection approach that ensures an inertial behavior of the weights over time.

Experiment 3: The third experiment aims to quantify the influence of context at different search steps. To this end, we experimented with Context and drop the use of the context at a given search round (e.g., at 1st, 4th, or 8th round). We then compare precision results with those obtained by Default (0-8). Figure 5 shows that if we switch off the contribution of the context at the very beginning of the search (round 1) we obtain a final average improvement of 55.55% over Default. The improvement increases when we keep using the context, obtaining a 66.66% improvement at round 4 and a 92.59% at round 8, respectively. This shows that not only the use of the context increases the search effectiveness, but also that the inertial behavior of the context switch provides even better results.

Visual Example: We conclude the section by showing actual query results obtained when using context. The aim is to provide evidence of the contribution of the context information with respect to the traditional approaches to similarity queries. In particular,

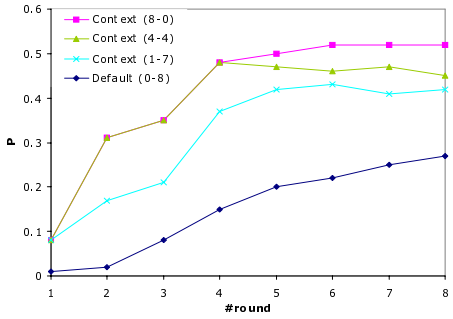


Fig. 5. Inertial behavior of the context measured at different rounds: Context vs Default.

in the example of Figure 6, showing the top 20 images for a query in the semantic class "animals in the forest", Context returns 5 relevant images out of 20 (25% precision), whereas Default does not return any relevant object.

7 Conclusions

In this paper we have presented an approach to contextualize image queries, which is able to effectively represent complex semantic concepts by means of the notion of image context. Although this is simple, it is indeed effective and does not require neither a-priori classification of the image database, nor the analysis of surrounding text (e.g., image caption, text of Web page including the image, etc.), which might not always be available with an image. Furthermore, our approach easily complements available relevance feedback techniques, representing a "good" starting point for interactive searches, and helps increasing both the effectiveness and efficiency of further rounds of retrieval. This has been demonstrated through experiments on a dataset of about 10,000 manually classified images.

In this paper, for sake of definiteness, we have analyzed the performance of our method by considering the case where rather simple feature descriptors are used. Further work will include a thorough investigation of the effects of using more complex features and distance functions for similarity assessment.

References

1. I. Bartolini, P. Ciaccia, and F. Waas. FeedbackBypass: A New Approach to Interactive Similarity Query Processing. In *Proceedings of the 27th International Conference on Very Large Data Bases (VLDB'01)*, pages 201–210, Rome, Italy, Sept. 2001.
2. H. Drucker, C. Burges, L. Kaufman, A. J. Smola, and V. Vapnik. Support Vector Regression Machines. In *Proceedings of the International Conference on Advances in Neural Information Processing Systems 9*, pages 155–161, 1997.
3. Y. Ishikawa, R. Subramanya, and C. Faloutsos. MindReader: Querying Databases Through Multiple Examples. In *Proceedings of the 24th International Conference on Very Large Data Bases (VLDB'98)*, pages 218–227, New York, NY, USA, Aug. 1998.
4. M. Ortega and S. Mehrotra. *Relevance Feedback in Multimedia Databases. In* Handbook of Video Databases: Design and Applications. CRC Press, 2003.

(a) (b)

Fig. 6. Visual results at the first round of search for the semantic concept "animals in the forest": Default (a) vs Context (b). Images in the same semantic class of the query are highlighted in green.

5. Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra. Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):644–655, 1998.
6. G. Salton. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Addison-Wesley, Reading, MA, 1989.
7. S. Santini and R. Jain. Integrated Browsing and Querying for Image Databases. *IEEE MultiMedia*, 7(3):26–39, 2000.
8. A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-Based Image Retrieval at the End of the Early Years. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(12):1349–1380, 2000.
9. J. R. Smith and S.-F. Chang. VisualSEEk: A Fully Automated Content-Based Image Query System. In *ACM Multimedia*, pages 87–98, Boston, MA, Nov. 1996.
10. S. Smoliar and L. Wilcox. Indexing the Content of Multimedia Documents. In *Proceedings of the Second International Conference on Visual Information Systems (VISual'97)*, pages 53–60, San Diego, CA, Dec. 1997.
11. S. L. Tanimoto. *An Iconic Symbolic Data Structuring Schema. In* Pattern Recognition and Artificial Intelligence. Academic Press, N.Y., 1976.
12. V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, N.Y., 1995.
13. H. Wang, B. C. Ooi, and A. K. H. Tung. iSearch: Mining Retrieval History for Content-Based Image Retrieval. In *Proceedings of the 8th International Conference on Database Systems for Advanced Applications (DASFAA'03)*, pages 275–283, 2003.
14. T. Westerveld. Image Retrieval: Content Versus Context. In *Proceedings of the Content-Based Multimedia Information Access, RIAO*, 2000.
15. L. Zhang, F. Qian, M. Li, and H. Zhang. An Efficient Memorization Scheme for Relevance Feedback in Image Retrieval. In *Proceedings of the International Conference on Multimedia & Expo (ICME'03)*, 2003.