SwiftLink: Serendipitous Navigation Strategy for Large-scale Document Collections

Marc von Wyl
Department of Computer Science
University of Geneva
Carouge, Switzerland
http://viper.unige.ch

Stéphane Marchand-Maillet
Department of Computer Science
University of Geneva
Carouge, Switzerland
Stephane.Marchand-Maillet@unige.ch

***Abstract*—The multiplication of large-scale document collections has created the need for robust and adaptive access strategies in many applicative areas. In this paper, we depart from the traditional *document search* paradigm to move onto the construction of a *collection navigation* strategy. We thus detail a model where user clicks are taken as expression of interest rather than positive search feedback. In terms of the interpretation of user interaction, our model is close to the recommendation paradigm. However, it exploits user feedback in a navigational procedure to perform an approximation of interactive metric learning procedures. Our model is directly compatible with distributed data indexing and is therefore inherently built to cope with scalability issues. The evaluation of such models is discussed and implemented over the context of an image collection. Experiments are presented where the user browses a collection of images with changing center of focus in the course of the browsing operations. Results show that our model is effectively able to recommend the user a relevant path through the collection and to adapt according to variable expressed interests.**

I. Motivation

The multiplication of large-scale document collections have created the need for robust and adaptive access strategies in many applicative areas such as education or culture [1], [2], [3], business [4] or public domain [5]. Search and retrieval have installed themselves as a base paradigm for accessing documents from within a collection. Alternative browsing systems have also been proposed, which are still mostly based on a search objective. Until now, systems have been oriented towards mental image localization [6], [7], copy detection [8] or iterative systems using user's relevance feedback [9], [10], [11].

It may however be the case that the user is not targeting a particular document or concept, but rather wishes to explore the collection in an opportunistic way. In that context, the user wishes to access "interesting" subsets of the collection. The challenge is therefore in defining, from the user's behavior, what part of the collection is of interest.

Such a challenge may be seen as a recommendation process where the user replaces the notion of a "query" (elsewhere also referred to as "mental image") with that of an "interest" (related with curiosity or judgement). The main difference is that *interest* may be less focused than a query and also consistent at a higher level. While relevant query responses are often assumed to be consistent in terms of the semantic concepts they encompass, it may not be the case here. Interest may be based on higher-level notions such as aesthetics or contrast.

Also in contrast to recommender systems, our approach does not require the definition of a user profile nor it imposes specific search sessions with pre-defined objectives. In other words, we present an efficient data access strategy enabling intuitive browsing by estimating the user's intention from his/her input to the system and displaying items that are considered as most interesting to him/her (and thus likely to be further accessed).

This paper reports our first steps into the definition of a document collection navigation strategy while keeping in mind the following objectives:

1) We wish to accommodate serendipity. We assume no pre-defined (fixed) objective of the user's chain of actions;
2) The system should fit classic (simple) interaction models;
3) The system should be scalable in terms of the size of the document collection.

In section II, we present our interaction model, which describes the type of interaction that is expected from the user and what information is carried over with this interaction. Section III formalises our navigation model, anticipating scalability issues. Our study includes an inherent temporal dimension, which makes the evaluation context different from that of classical search systems. Section IV proposes to evolve classical evaluation models to accommodate such difference.

II. Method

Our aim is to construct a system that sees user interaction as an *expression of interest* and uses this to provide further recommendations to the user. While social recommendation [12] propagates user interests across a community, here, recommendation is based on the navigation history, in the line of Adaptive Hypermedia Navigation [13].

The main features of our model are the use of positive only feedback and a user model with time-limited characteristics. The main idea is to discover, at every step of the navigation, what linear combination of visual features brings together past choices of the user and to adapt accordingly in a scalable fashion. We propose the following basic navigation scenario that may be implemented over essentially any type of visualization device. The main control at every iteration is a single click on a document. Here, clicking a document means mining deeper in the direction of that document, in relation to a past navigation behavior. While allowing specific target search, this process allows for widening or reorienting the navigation by simply breaking the coherence in the history of actions.

1) The system samples the collection from the best of its current knowledge;
2) The user activates one of the proposed documents;
3) The system updates its knowledge over the session accordingly and repeats from Step 1.

In the sequel, we detail how we model step 1 and 3 (system adaptation), according to what is assumed from step 2 (user input). Essentially, we are based on the setup of metric learning [14] to adapt a measure over the collection and to be able to extract a collection sample suited for further navigation.

This paper focuses on the formal back-end model for collection navigation and document recommendation. The issue of the display and how the interaction is captured, depending on the used device is the subject of future work.

III. Navigating document collections

A. Model

The user is faced with a collection of N documents $\mathcal{C} = \{x_1, \cdots, x_N\}$, represented into a M-dimensional feature space ($x_i = [x_{i1} \cdots x_{iM}]^\mathsf{T}$). Given $A \in \mathcal{R}^{M\times M}$, a symmetric positive semi-definite matrix, we use the Mahalanobis distance between documents

$$d_A^2(x_i, x_j) = (x_i - x_j)^\mathsf{T} A (x_i - x_j) \quad (1)$$

As posed in the metric learning setup [14], we wish to use matrix A as a way of adapting to user feedback.

Typically, by iterating document selection (step 2), the user generates a navigation history in the form of an ordered document subset

$$\mathcal{H}_t = \{x^{(1)}, x^{(2)}, \cdots, x^{(t)}\}$$

where $x^{(t)} \in \mathcal{C}$ is the document selected at iteration t. By learning a scaling matrix A fit to the navigation history (step 3), the system will then adapt its base distance $d_A(.,.)$ to group documents $x^{(t)}$ close together according to a defined policy (detailed below). From there on, the system will then sample the collection (step 1) by ranking documents by their distance d_A from the latest chosen sample $x^{(t)}$ and displaying top-ranked documents. The initial collection sample may arrive from the top-ranked responses to a query or a random sample of the collection.

By changing the distance function in Equation (1), we apply a linear transform on the feature space. If $V \in \mathcal{R}^{M\times M}$ is such that $A = V^\mathsf{T} V$, then $d_A(x_i, x_j) = d_{\mathrm{Id}}(y_i, y_j)$, where $y_i = V x_i$ and Id is the identity matrix. In particular, if A is taken as Σ^{-1}, the inverse of the $M \times M$ feature covariance matrix computed over $\mathcal{C}$, then the eigendecomposition $A = U^\mathsf{T} \Lambda U$ allows for defining decorrelated features y_i (the same notations will then apply for y_i, $y^{(t)}$ and y_{ik}).

We now use this property to simplify the computation of our new distance function and make it scalable (see section III-B). We first generalize by using $d_{B^{(t)}}(y_i, y_j)$ as distance function with $B^{(t)} = \mathrm{diag}(1/\sigma_1^{(t)2}, \cdots, 1/\sigma_M^{(t)2})$ a diagonal matrix, initially set to $B^{(0)}$ =Id, the identity matrix (ie $\sigma_k^{(0)} = 1\ \forall k$).

At step t, to evaluate whether two documents $y^{(t-1)}$ and $y^{(t)}$ are close to each other in terms of feature k, we compare their distance with the corresponding scaling factor $\sigma_k^{(t)}$. If $|y_{ik}^{(t)} - y_{jk}^{(t-1)}| < \sigma_k^{(t)}$, we assume a consistent navigation behavior and feature k should be reinforced to keep the documents close to each other. We thus update $\sigma_k^{(t)}$ following:

$$\sigma_k^{(t)} = \lambda_t \sigma_k^{(t-1)} + (1 - \lambda_t)|y_{ik}^{(t)} - y_{jk}^{(t-1)}| \quad (2)$$

where $\lambda_t \in [0, 1]$ is the strength of contraction.

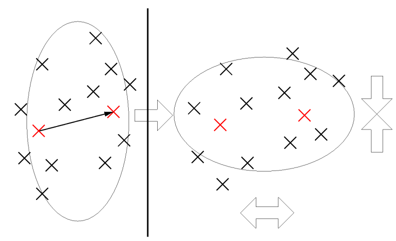


Fig. 1. Space modification after two consecutive choices from the user

Figure 1 illustrates the idea of contraction of the feature space on the vertical feature and expansion on the horizontal feature.

The choice of λ_t is further automated by taking into account the temporal value of the user's decisions. We introduce a forgetting framework that gives more weight to recent choices. We weight the past choices using a sigmoid function $S(\tau) = \frac{1}{1+e^{-\tau}}$ over interval $\tau \in [-a, +a]$ and weight $\mathcal{H}_t$ on this function over a fixed number of steps $t, \ldots, t - N_h$.

Hence at iteration t, we compute weights $\lambda_t, \cdots, \lambda_{t-N_h}$ using the weighting function $S(.)$ with $\tau = (t-1)2a/(N_h-1)$. In summary, with proper normalization, we obtain the following update rule for $\sigma_k^{(t)}$.

$$\sigma_k^{(t)} = \sum_{l=0}^{N_h-1} \left(S\left(l \frac{2a}{(N_h - 1)} \right) \frac{|y_{ik}^{(t-l)} - y_{jk}^{(t-l-1)}|}{\sigma_k^{(t-l-1)}} \right) \quad (3)$$

A part of the evaluation focuses on testing the influence of the choice of a value for N_h, as described next.

B. Scalability

The above model is a simplification of typical distance metric learning algorithms [14]. A first difference with the usual context within metric learning applies is that, in our context, we do not look for classification or clustering, whether or not based on constraints [15], [16].

Our model is also designed to cope with scalability. Basically, we secure the possibility for most computations to be done offline and then efficiently organise for online adaptation.

The modification of matrix $B^{(t)}$ is straightforward using Equation (3) ($O(M)$). The computation of distance $d_{B^{(t)}}(.,)$ requires the computation of the sum in

$$d^2_{B^{(t)}}(y_i, y_j) = \sum_{k=1}^{M} (\frac{1}{\sigma_k^{(t)}}(y_{ik} - y_{jk}))^2$$

and values of $(y_{ik} - y_{jk})$ are fixed and may be properly indexed prior to starting the process. Large-scale distributed indexing strategies exist for resolving the nearest neighbour ranking problem [17]. The use of such indexing strategies in every isolated feature space (ie $(y_{ik} - y_{jk})$) to prepare the data is not addressed here and is the subject of current and future work [18].

IV. Evaluation Strategy and Experiments

A. Dataset

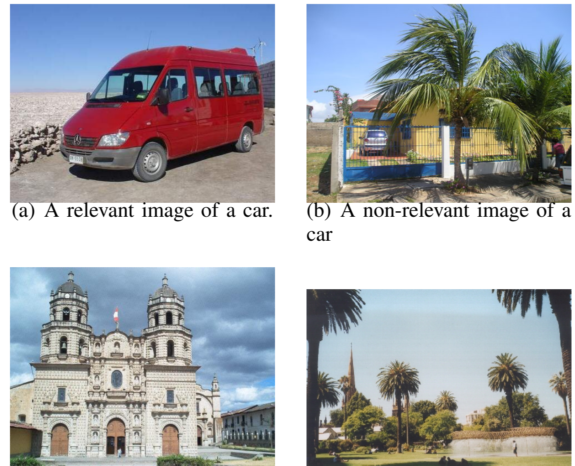
We consider images as documents here but our model readily applies on all types of documents, using appropriate features. As initial step, we use a small subset of the IAPR-TC12 dataset [19][1]. The images are manually annotated based on the names of the elements within the image. Tags are hidden to the user and we assume that the user selection will be based on the occurrence of a given element (related to a tag) in the image. We will evaluate precision a posteriori using image tags. In order to make the evaluation less ambiguous, we filter the annotation by a pseudo-manual process. Given a word in the annotation vocabulary, from the set of images tagged with that word, we select those images where the element appears clearly in the image. For example, some images are annotated with the word "car" even if a car barely appears on it. A user looking for relevant images of a car will not consider those as relevant. As a further motivation, Figure 2 shows an example of relevant (selected) and non-relevant (discarded) images regarding example tags.

We used this procedure to create several benchmarks with the words "car" (82 documents), "group" (51 documents), "castle or church" (235 documents), "portrait" (82 elements).

B. Document representation

Here, we seek a feature space based on visual features but less specific than that of a copy detection system like SIFT[8] features. We choose to use a more general image representation based on the MPEG7 framework [21] where images are described by three different aspects : the colors, the shapes and the texture. Hence, we use the Grid Color Moment feature with 81 dimensions [22], the Edge Direction Histogram with 16 dimensions [23] and the Gabor Texture Moment with 24 dimensions [24] features. These features have already prove they usefulness in our previous work [10]. Every image is therefore represented by a vector of $M = 121$ components,

[1]We are currently running large-scale experiments over the large-scale annotated image collection ImageNet [20].



(a) A relevant image of a car. (b) A non-relevant image of a car (c) A relevant image of a church. (d) A non-relevant image of a church

Fig. 2. Relevant (a,c) and non-relevant (b,d) images annotated with the words "car" and "church", respectively

that we normalize by groups of features and compute the covariance matrix over our collection to obtain an empirically normalized space (using the usual hypothesis of a Gaussian distribution of features).

C. Strategy and Results

a) Document recommendation: We first evaluate the scenario where the user concentrates on one specific topic and digs into it for a number of iterations. Hence, we randomly select an image as seed example and use it as a query. We then select, at every iteration, the first relevant image in the ranking that was not chosen yet in the navigation history. We evaluate the performance using Mean Average Precision (MAP):

$$MAP = \frac{\sum_{q=1}^{Q} AveP(q)}{Q} \quad (4)$$

where $AveP$ is the average precision

$$AveP = \frac{\sum_{k=1}^{N} (P(k) \times rel(k))}{\text{number of relevant documents}} \quad (5)$$

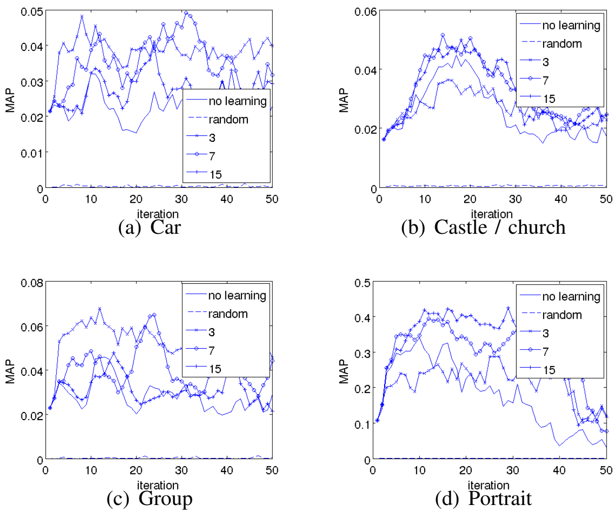
Q is the number of queries over which the average is proposed, and N is the size of the collection, $rel(k)$ equals 1 if the image k in the ranking is relevant, 0 otherwise. $P(k)$ is the precision at position k

$$P(k) = \frac{\text{number of relevant documents at index } k}{k}. \quad (6)$$

Results are averaged over 20 experiments.

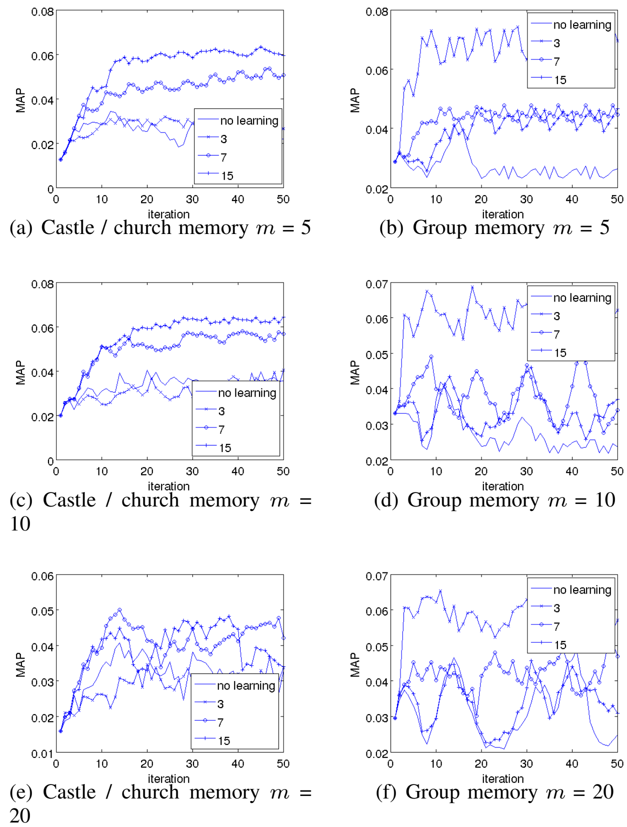
We compare our strategy for several values of N_h with a system with no learning and a monkey system generating random ranking. Figure 3 shows that for all classes, we always improve over a no-learning system. The actual choice for an optimal value of N_h seems sensitive to the navigation objective.

The instability in the results are analyzed to originate from the way the features group the documents. When data is



(a) Car (b) Castle / church (c) Group (d) Portrait

Fig. 3. Mean Average Precision



(a) Castle / church memory m = 5 (b) Group memory m = 5 (c) Castle / church memory m = 10 (d) Group memory m = 10 (e) Castle / church memory m = 20 (f) Group memory m = 20

Fig. 4. Retrieval MAP. m is the number of iterations before an items may be selected again. Smaller values of m make learning more stable due to the repetition of examples.

clustered in the feature space, the navigation strategy is stable in visiting one cluster and should re-adapt to visit another class cluster. This generating instability in the MAP curves.

To cater for this, we adapt our evaluation setup, modeling a user with limited memory. Instead of selecting an element if it was never selected before, we select it if it was never selected in the m previous steps, allowing some reinforcement by looping over documents. Figure 4 shows the results with different sizes of memory m on the sets “castle or church” and “group”.

We see that users benefit from re-selecting images in their navigation strategy. Doing so, the system reinforces related subgroups and aggregates them in a stronger manner, which supports our hypothesis on the existence of small clusters in the first part of this evaluation.

b) Opportunistic browsing: The second part of the evaluation concentrates on changes in the behavior of the user. More specifically, we wish to evaluate how fast the system can react to a change in the navigation objective. This is important as it is precisely this capability that distinguishes our system from a target search system.

We model a user who shows a consistent behavior for a number of iterations (as above) and then switches to another topic. This is modeled as the above process apart from the fact that at a given point, the model switches to a randomly chosen topic by selecting a random image in the display. We stick to our limited-memory user, as above, with $m = 5$ and make a topic transition after 20 iterations. Figure 5, shows related results for several class combinations.

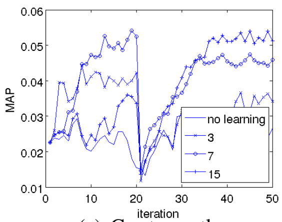
We see that our strategy has no problem adapting to the new class. The transition is illustrated by an abrupt drop in the MAP values but the system pick up quickly on the new topic and reaches again a larger MAP value than the no-learning system.

V. Conclusion

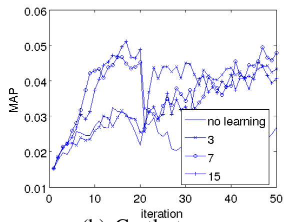
We show preliminary results in the construction of a navigation system, where the interaction is simplified to a click-through operation. This elementary feedback, interpreted as an expression of interest, is used to learn an approximate metric that groups elements selected in the navigation history. This enables the system to act as a local recommender for the user, based on the current navigation history, modulated with a forgetting factor. The context of metric learning is used to justify approximations.

We have constructed an evaluation strategy based on image annotation and oriented towards MAP values. Scalabilty is embedded in our model and we are running experiments to test for this aspect. Our evaluation is biased towards the idea of retrieval and we plan to complete it with actual user experiments to evaluate the value of an associate tool in discovering collection content. Such an evaluation would also bring in the picture the question on the form of display to present to the user, which was not addressed here.

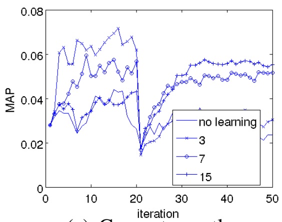
There are many evidences on the utility of navigation-based systems, in complement to search and retrieval operations. However, formal modeling and evaluation of such systems is still not mature enough. We plan to further develop our methodology with applications in the educational sector where



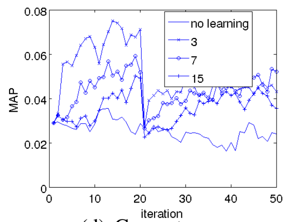
(a) Car to castle



(b) Castle to car



(c) Group to castle



(d) Group to car

Fig. 5. Retrieval MAP obtained when the user browses a topic for 20 iterations and then switches focus of interest (memory m = 5). For suitable values of latency N_h in the model, our system recovers quickly from the change.

the imperfection of document representations may become an advantage [1], [2], [3].

VI. Acknowledgments

This work has been partly funded by SNF fund No. 200021-119906, in parallel with the Swiss NCCR(IM)2 and the ICT COST Action IC1002 “Multilingual and Multifaceted Interactive Information Access” (MUMIA) .

References

[1] P. Brusilovsky, L. N. Cassel, L. M. Delcambre, E. A. Fox, R. Furuta, D. D. Garcia, F. M. S. III, and M. Yudelson, “Social navigation for educational digital libraries,” *Procedia Computer Science*, vol. 1, no. 2, pp. 2889 – 2897, 2010.

[2] A. Carbonaro, “Collaborative and semantic information retrieval for technology-enhanced learning,” in *Proceedings of the 3rd International Workshop on Social Information Retrieval for Technology-Enhanced Learning (SIRTEL’09)*, Aachen, Germany, 2009.

[3] N. Manouselis, H. Drachsler, R. Vuorikari, H. G. K. Hummel, and R. Koper, *Recommender Systems Handbook*. Berlin: Springer, 2011, ch. Recommender Systems in Technology Enhanced Learning, pp. 387–415, (P. B. Kantor, F. Ricci, L. Rokach, and B. Shapira (Eds.)).

[4] S. Seidel, F. M. Muller-Wienbergen, M. Rosemann, and J. Becker, “A conceptual framework for information retrieval to support creativity in business processes,” in *Proceedings 16th EuropeanConference on Information Systems*, Galway, Ireland, 2008.

[5] R. van Zwol and B. Sigurbjörnsson, “Faceted exploration of image search results,” in *Proceedings of the 19th international conference on World Wide Web (WWW’2010)*, 2010.

[6] M. Ferecatu and D. Geman, “A statistical framework for image category search from a mental picture,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 6, pp. 1087–1101, jun 2009.

[7] I. J. Cox, M. L. Miller, T. P. Minka, T. V. Papathomas, and P. N. Yianilos, “The bayesian image retrieval system, pichunter: Theory, implementation and psychophysical experiments,” *IEEE transactions on image processing*, vol. 9, pp. 20–37, 2000.

[8] J. Sivic and A. Zisserman, “Video google: a text retrieval approach to object matching in videos,” in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, oct. 2003, pp. 1470 –1477 vol.2.

[9] E. Bruno, J. Kludas, and S. Marchand-Maillet, “Combining multimodal preferences for multimedia information retrieval,” in *Proceedings of the international workshop on Workshop on multimedia information retrieval*, 2007.

[10] M. von Wyl, H. Mohamed, E. Bruno, and S. Marchand-Maillet, “A parallel cross-modal search engine over large-scale multimedia collections with interactive relevance feedback,” in *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*, ser. ICMR ’11. New York, NY, USA: ACM, 2011, pp. 73:1–73:2.

[11] F. Yan, K. Mikolajczyk, and J. Kittler, “Multiple kernel learning via distance metric learning for interactive image retrieval,” in *Proceedings of the 10th international conference on Multiple classifier systems*, ser. MCS’11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 147–156.

[12] D. H. Park, H. K. Kim, I. Y. Choi, and J. K. Kim, “A literature review and classification of recommender systems research,” *Expert Systems with Applications*, 2012, (in press).

[13] P. Brusilovsky, “Adaptive hypermedia,” *User Modelling and User-Adapted Interaction*, vol. 11, pp. 87–110, 2001.

[14] B. McFee and G. Lanckriet, “Metric learning to rank,” in *Proceedings of the 27th International Conference on Machine Learning (ICML’10)*, June 2010.

[15] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall, “Learning a mahalanobis metric from equivalence constraints,” *Journal of Machine Learning Research*, vol. 6, pp. 937–965, 2005.

[16] M. Bilenko, S. Basu, and R. J. Mooney, “Integrating constraints and metric learning in semi-supervised clustering,” in *Proceedings of 21st International Conference on Machine Learning (ICML-2004)*, Banff, Canada, July 2004, pp. 81–88. [Online]. Available: http://www.cs.utexas.edu/users/ai-lab/pub-view.php?PubID=51547

[17] G. Amato and P. Savino, “Approximate similarity search in metric spaces using inverted files,” in *Proceedings of the 3rd international conference on Scalable information systems*, ser. InfoScale ’08. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008, pp. 28:1–28:10. [Online]. Available: http://dl.acm.org/citation.cfm?id=1459693.1459731

[18] H. Mohamed and S. Marchand-Maillet, “Distributed media indexing based on mpi and mapreduce,” in *10th Workshop on Content-Based Multimedia Indexing (CBMI 2012)*.

[19] M. Grubinger, P. Clough, H. Müller, and T. Deselaers, “The IAPR benchmark: A new evaluation resource for visual information systems,” in *International Conference on Language Resources and Evaluation*, Genoa, Italy, May 2006.

[20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” in *CVPR09*, 2009.

[21] A. P. Shih-Fu Chang, Thomas Sikora, in *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 11, june 2001, pp. 688–695.

[22] M. O. M. Stricker, “Similarity of color image,” *Proc. SPIE*, vol. 2420, pp. 381 – 392, 1995.

[23] D. B. M. Swain, “Color indexing,” *International Journal of Computer Vision*, vol. 7(1), pp. 11–32, 1991.

[24] A. K. Jain and A. Vailaya, “Image retrieval using color and shape,” *Pattern Recognition*, vol. 29, pp. 1233 – 1244, 1996.