www.ietdl.org

Published in IET Image Processing
Received on 19th October 2012
Accepted on 21st November 2013
doi: 10.1049/iet-ipr.2013.0514

ISSN 1751-9659

Query-dependent metric learning for adaptive, content-based image browsing and retrieval

Junwei Han[1], Stephen J. McKenna[2]

[1]School of Automation, Northwestern Polytechnical University, People's Republic of China
[2]School of Computing, University of Dundee, Dundee DD1 4HN, UK
E-mail: junweihan2010@gmail.com

Abstract: Content-based image retrieval (CBIR) systems often incorporate a relevance feedback mechanism in which retrieval is adapted based on users identifying images as relevant or irrelevant. Such relevance decisions are often assumed to be category-based. However, forcing a user to decide upon category membership of an image, even when unfamiliar with a database and irrespective of context, is restrictive. An alternative is to obtain user feedback in the form of relative similarity judgments. The ability of a user to provide meaningful feedback depends on the interface that displays retrieved images and facilitates the feedback. Similarity-based 2D layouts provide context and can enable more efficient visual search. Motivated by these observations, this study describes and evaluates an interactive image browsing and retrieval approach based on relative similarity feedback obtained from 2D image layouts. It incorporates online maximal-margin learning to adapt the image similarity metric used to perform retrieval. A user starts a session by browsing a collection of images displayed in a 2D layout. He/she may choose a query image perceived to be similar to the envisioned target image. A set of images similar to the query are then returned. The user can then provide relational feedback and/or update the query image to obtain a new set of images. Algorithms for CBIR are often characterised empirically by simulating usage based on pre-defined, fixed category labels, deeming retrieved results as relevant if they share a category label with the query. In contrast, the purpose of the system in this study is to enable browsing and retrieval without predefined categories. Therefore evaluation is performed in a target-based setting by quantifying the efficiency with which target images are retrieved given initial queries.

1 Introduction

Most early content-based image retrieval (CBIR) systems relied on pre-defined image-to-image similarity measures. These so-called computer-centric systems were relatively easy to implement but inherent drawbacks limited performance. Image understanding is highly subjective; each user has different intentions and preferences when searching for images or browsing an image collection. These can vary from session to session even if identical queries are posed initially. A user might decide to choose a new query during a session because what they seek has changed, perhaps as a result of inspiration gained from browsing.

A great deal of effort has been devoted to developing category-based image retrieval. Although some works [1–3] attempt to solve category image search from the viewpoint of generating powerful features with the ability to discriminate images across categories, many also incorporate relevance feedback (RF) techniques into the systems. RF was proposed to address the limitation of using fixed similarity measures. In RF, online learning of query-dependent similarity measures is performed based on user feedback. When adopting this approach it is often assumed that a user is looking for a category of images and starts with a query from that category. The user provides feedback on the categorical relevance of retrieved images. Machine learning methods such as support vector machines [4], manifold learning [5], biased maximum margin analysis [6] or manifold ranking [7], are used to train the classifier or refine the similarity measure online based on the binary labels. Assuming that other users may have similar interests, another group of methods [8–11] accumulates logs of previous users' relevance judgments and performs offline learning. The offline learned information can be integrated with the current user's online RF for the collaborative learning of similarity metrics.

Requiring feedback in terms of category-membership decisions burdens users by forcing them to decide upon a useful categorisation of images even if unfamiliar with the database [12]. Although this can be appropriate for a category-based search, it is not so appropriate for browsing and retrieval more generally. An example of an alternative approach is the target-based retrieval system proposed by Cox *et al.* [12] in which a probability distribution over possible targets is iteratively updated based on user behaviour. The probability of an image being the target is increased if it is closer to relevant examples in feature space.

RF in CBIR is usually binary; users identify a set of examples as 'relevant' and identify the others as 'irrelevant' [4–11]. However, it is often hard and unnatural for users to make these binary decisions because they may have a large variety of ways to understand images. Suppose a user

Fig. 1 *Binary RF is restrictive*

a Image of a fashion model
b Image that might be of a fashion model
c Image that is definitely not of a fashion model

intends to search for images of fashion models. The user can easily identify Fig. 1*a* as relevant and Fig. 1*c* as irrelevant. It is not clear, however, whether or not Fig. 1*b* is of a fashion model. For such a case, qualitative and relative comparison of images is meaningful and preferred. We could all agree on the relative interpretation that Fig. 1*a* is more likely to be of a fashion model than Fig. 1*b* and Fig. 1*b* is more likely to be of a fashion model than Fig. 1*c*. Indeed, RF based on relative information can provide consistent responses from users; it can thus improve the precision of communication between users and CBIR systems. This paper proposes to build a model of relative RF.

The study of how to learn from relative comparisons is attracting increasing attention. Joachims [13] proposed a ranking-SVM method that converted the learning task to standard SVM classification. Schultz and Joachims [14] extended this approach to learn distance metrics. Freund *et al.* [15] presented RankBoost. Rank-based distance learning has been used to solve vision problems. Frome *et al.* [16] proposed a method to learn local image-to-image distance functions for classification based on relative comparisons. Hu *et al.* [17] explored a multiple-instance ranking approach based on ranking-SVM to order images within each category for retrieval. Lee *et al.* [18] employed a rank-based distance metric to retrieve images of tattoos. Faria *et al.* [19] also used rank learning for CBIR. Huang *et al.* [20] applied metric learning from rank correlations to a medical image retrieval. Wang *et al.* [21] proposed a regularised kernel machine to use comparative object similarity for object recognition. Recently, Parikh *et al.* [22] presented a novel methodology for modelling relative visual attributes. It learned a ranking function which can predict the strength of an attribute in an image with respect to other images. Experiments demonstrated that relative attribute prediction had advantages over binary attribute prediction. These methods [16–22] were investigated under the scenario of offline learning.

The ability of a user to provide meaningful feedback depends not only on the quality of retrieval but also on the manner in which the interface displays retrieved images and facilitates user feedback. A well-designed interface will make this interaction easy and enjoyable for users, and enhance efficiency. Rodden [23] performed user studies that demonstrated that 2D layouts could enable users to find a target image or group of images more quickly than when images were arranged in order of similarity to a query image. Moghaddam *et al.* [24] also argued that visualising images in a 2D space can be superior, allowing mutual similarities to be reflected. Wang *et al.* [25] proposed the high-entropy layout distributions (HELD) method for arranging image collections and this method is adopted here for display of images in a way that facilitates browsing and feedback.

This paper describes an interactive image browsing approach that incorporates online, rank-based learning using relative information into an efficient interface that visualises image collections in 2D space and facilitates users to provide relational feedback. In this scenario, a user intends to find his/her envisioned target images. For example, a fashion designer may seek a particular kind of fabric or a customer may look for a particular style of shoe online. The proposed system provides users with target-based image retrieval that is different from traditional category-based retrieval. To achieve this end, we develop an architecture that seamlessly combines image browsing, image retrieval and RF. A user starts a session by browsing a collection of images displayed in a 2D layout. He/she may choose a query image perceived to be similar to the envisioned target image. A set of images similar to the query are returned. The user can offer relational feedback and update the query image to refine results. This relies on a ranking based learning of image relationships.

The proposed work builds on an earlier conference paper [26] and makes the following contributions. A user feedback mechanism is proposed based on HELD displays in which the user updates the query and can select and arrange relevant images in relation to it, generating relative feedback. An experimental evaluation is presented based on user simulations that quantify performance with respect to the method's parameters and in comparison to alternative strategies. It should be noted that algorithms for CBIR are often characterised experimentally by simulating usage based on pre-defined, fixed category labels, deeming retrieved results to be relevant when they share a category label with the query. In contrast, the purpose of the system in this paper is to browse in a non-categorical manner. Therefore a different evaluation method is proposed.

The rest of this paper is organised as follows. Section 2 describes the framework, Section 3 presents results and conclusions are drawn in Section 4.

2 Browsing and retrieval framework

2.1 Overview

The iterative browsing framework is summarised in Algorithm 1 (see Fig. 2). Given an image collection, I, an initial query image, q_0 and an initial distance metric, D_0, (typically Euclidean), the N closest matches to the query are retrieved (Section 2.2) and displayed using an automatically generated 2D layout, L_0 (Section 2.3). Note that if an initial query was not available, an initial layout could be generated based on a representative (or randomly selected) subset of I of cardinality N. A browsing session then consists of a sequence of iterations that continue until the user decides to end the session. For example, if the user had a target image in mind, the session ends once this image is found. At the tth iteration, the user is presented with a 2D image layout L_{t-1}. The user selects a query image, q_t, from this layout. This query image may or may not differ from the previous query, q_{t-1}. The user selects further relevant images from the layout and rearranges these on the display to qualitatively express judgments about their relative similarities to the query. The selected images and their implied ordering are used to generate a set of inequality constraints, P_t, automatically (Section 2.4). A learning algorithm then uses the selected query and the constraints to obtain a new metric, D_t (Section 2.5). This metric is used to retrieve the closest matches from the image set and a visualisation algorithm produces a new 2D layout from these matches.

Algorithm 1

$t = 0$

$\{I, q_0, D_0\} \xrightarrow{\text{matcher and visualiser}} \{L_0\}$

repeat

  $t = t + 1$

  $\{L_{t-1}\} \xrightarrow{\text{user}} \{q_t, P_t\}$

  $\{I, q_t, P_t\} \xrightarrow{\text{learner}} \{D_t\}$

  $\{I, q_t, D_t\} \xrightarrow{\text{matcher and visualiser}} \{L_t\}$

Until user ends session

Fig. 2 *Interactive browsing*

2.2 Retrieving images

The problem of retrieving N images that are closest to a query q based on high-dimensional feature vectors has been studied extensively. We only note here that efficient algorithms exist for performing these nearest neighbour searches approximately (e.g. [27]). Let $\boldsymbol{q}$ and $\boldsymbol{x}$ denote feature vectors representing a query and a database image, respectively. The dissimilarity of these images can be measured using a parameterised Mahalanobis distance $D(\boldsymbol{q}, \boldsymbol{x}; \boldsymbol{W}) = \sqrt{(\boldsymbol{q}-\boldsymbol{x})^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{q}-\boldsymbol{x})}$. The symmetric matrix $\boldsymbol{W}$ is positive semi-definite to ensure that D is a valid metric, that is, $\boldsymbol{W} \geq 0$. If $\boldsymbol{W}$ is diagonal, the metric becomes a weighted Euclidean distance

$$D(\boldsymbol{q}, \boldsymbol{x}; \boldsymbol{w}) = \sqrt{\langle \boldsymbol{w} \cdot ((\boldsymbol{q}-\boldsymbol{x}) * (\boldsymbol{q}-\boldsymbol{x}))\rangle} \tag{1}$$

where '$\langle\cdot\rangle$' denotes inner product and '*' the element-wise product of two vectors. $\boldsymbol{w}$ is the vector consisting of diagonal elements of $\boldsymbol{W}$. In this case, $\boldsymbol{w} \geq 0$. It is this parametric family of diagonal metrics that is used in the remainder of this paper.

2.3 Generating 2D layouts

At each iteration of browsing, retrieved images are arranged automatically for display using HELD, a 2D layout generation algorithm [25]. Fig. 3 shows an example layout of 50 images retrieved based on colour correlogram features [1]; the query image is also shown in the upper left corner. The algorithm arranges images such that the amount of image overlapping is low and images with similar content appear close together. HELD generates layouts that conform

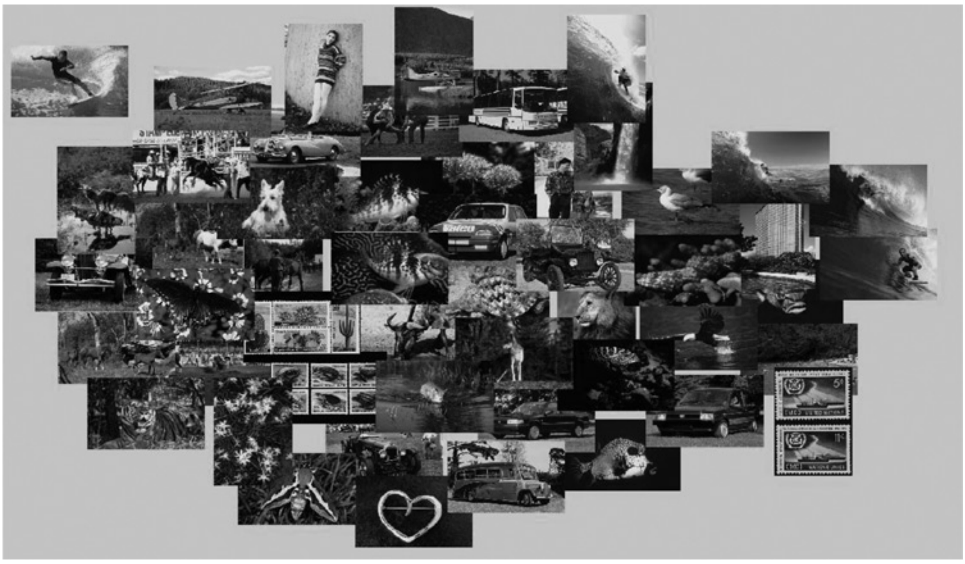

Fig. 3 *50-image HELD layout generated using MDS. The query is shown in the upper left corner*

doi: 10.1049/iet-ipr.2013.0514

to the shape of the available display region, approximate the high-dimensional image feature distribution, and make good use of the available layout space. HELD achieves this trade-off by optimisation of an objective function that combines dimensionality reduction with a layout entropy measure. The image set is used to define a density function in the 2D layout space and distributions with low differential Renyi entropy are penalised since they result in layouts in which some regions are over-populated and other regions are sparsely populated. In [20], the ISOMAP algorithm was used within HELD. Here instead multi-dimensional scaling (MDS) was used because of the relatively small number of images per layout. The distance in feature space between two images was measured using the current browsing metric, D_t.

2.4 Obtaining user feedback

The user might find what they seek in the current layout and decide to terminate the session. If, on the other hand, they decide to continue they have the option to select one of the newly retrieved images as a new query. A user might decide to choose a new query because it matches more closely what they seek or because what they seek has changed, perhaps as a result of inspiration gained. The user can then select a subset of images on the layout as being of relevance. Once this selection has been made the other images are removed from the display. Fig. 4*a* shows an example in which six images have been selected from a layout as relevant.

The user can then provide qualitative feedback by arranging the selected images to reflect their perceived relative similarity to the query. Images that are similar to the query are placed close to it and images that are less similar are placed further away. Fig. 4*b* shows the example in Fig. 4*a* after the user has arranged the images. This user-defined layout yields an ordering in terms of similarity to the query. This ordering implies a set of inequalities on the distance measure being employed by the user.

If the user arranges M images relative to the query then there are generally $M(M-1)/2$ inequalities expressing order relationships between these M images. If we assume the user's measure is a metric then most of these relationships are redundant and only $M-1$ inequalities are needed. In the example shown in Fig. 4*b* the constraints would be

$$\begin{aligned} P_t = \{ & (D_t(q_t, 1; \boldsymbol{w}) < D_t(q_t, 5; \boldsymbol{w}), D_t(q_t, 5; \boldsymbol{w}) \\ & < D_t(q_t, 4; \boldsymbol{w}), D_t(q_t, 4; \boldsymbol{w}) \\ & < D_t(q_t, 3; \boldsymbol{w}), D_t(q_t, 3; \boldsymbol{w}) < D_t(q_t, 6; \boldsymbol{w}), D_t(q_t, 6; \boldsymbol{w}) \\ & < D_t(q_t, 2; \boldsymbol{w})\} \end{aligned}$$

Additionally, non-selection of an image by the user can be taken to imply that it is more dissimilar to the query than any selected image. In this case, there are an additional $N-M$ inequalities giving a total of $N-1$ constraints.

2.5 Adapting the metric

The query and the user-generated constraints provide feedback information about the perceptual measures currently being employed by the user. A learning algorithm is used to determine a new distance metric that makes use of this information. Specifically, the objective of the learner is to infer the parameter $\boldsymbol{w}_t$ of the distance metric $D_t(.,.;\boldsymbol{w}_t)$. Ideally, this metric should satisfy the constraints P_t. (Henceforth, the subscript t is omitted). This learning task can be performed using a maximal-margin formulation with slack variables amounting to solving the following optimisation problem which has the same form as in [14, 16].

$$\begin{aligned} & \min_{\boldsymbol{w},\xi_{(q,i,j)}} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum\nolimits_{(q,i,j)} \xi_{(q,i,j)} \\ & s.t. \\ & \forall(D(q, i; \boldsymbol{w}) > D(q, j; \boldsymbol{w})) \in P: \\ & D^2(q, i; \boldsymbol{w}) - D^2(q, j; \boldsymbol{w}) \geq 1 - \xi_{(q,i,j)} \\ & \forall(q, i, j): \xi_{(q,i,j)} \geq 0 \\ & \boldsymbol{w} \geq 0 \end{aligned} \tag{2}$$

The term $\|\boldsymbol{w}\|^2$ measures structural loss, $\xi_{(q,i,j)}$ are slack variables and C is a trade-off parameter. Substituting (1) into the first set of constraints in (2) yields

$$\langle \boldsymbol{w} \cdot (\boldsymbol{d}_{q,i} - \boldsymbol{d}_{q,j}) \rangle \geq 1 - \xi_{(q,i,j)} \tag{3}$$

where $\boldsymbol{d}_{q,i} = (\boldsymbol{q}-\boldsymbol{x}_i)*(\boldsymbol{q}-\boldsymbol{x}_i)$ and $\boldsymbol{x}_i$ is the feature vector for the ith image. The final constraint $\boldsymbol{w} \geq 0$ ensures the learned distance is a metric. Without this constraint, the optimisation setting is the same as that of ranking-SVM and standard quadratic programming solvers such as SVM-Light could be used [13]. Ranking-SVM learns a ranking function to sort the data; elements of the parameter $\boldsymbol{w}$ can be negative thus the ranking values can be negative. Although image retrieval can be formulated as a ranking problem

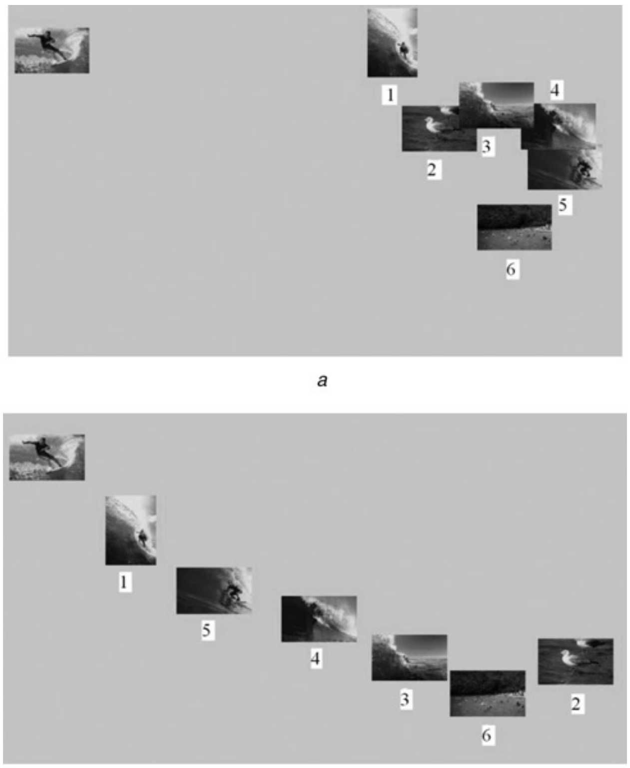

Fig. 4 *Example of user feedback*

a Six images selected as relevant in addition to a query image
b User has arranged these images reflecting their perceived similarity to the query

using such an approach [17], this is unsuitable for query-by-example. This is because ranking-SVM can have negative outputs, implying that some images are more similar to the query than the query itself (since the output for the query will be zero). This point is demonstrated empirically in Section 3.

Frome [16] proposed a custom dual solver that can guarantee non-negativity of $\boldsymbol{w}$. It is fast enough to be suitable for online learning during real-time browsing. This solver iteratively updates dual variables until convergence

$$\boldsymbol{w}^{(t)} = \max\left\{\sum_{(q,i,j)} \alpha^{(t)}_{(q,i,j)}\left(\boldsymbol{d}_{q,i} - \boldsymbol{d}_{q,j}\right), 0\right\} \quad (4)$$

$$\alpha^{(t+1)}_{(q,i,j)} = \min\left\{\max\left\{\frac{1 - \langle \boldsymbol{w}^{(t)} \cdot \left(\boldsymbol{d}_{q,i} - \boldsymbol{d}_{q,j}\right)\rangle}{\left\|\boldsymbol{d}_{q,i} - \boldsymbol{d}_{q,j}\right\|^2} + \alpha^{(t)}_{(q,i,j)}, 0\right\}, C\right\} \quad (5)$$

where $0 \le \alpha_{(q,i,j)} \le C$ are dual variables initialised to zero, see [16] for implementation details.

3 Evaluation

3.1 Data set

10,009 images from the Corel dataset were used including at least 100 images from each of 79 categories with semantic labels such as tiger, model and castle. These labels were used only to ensure a varied data set. It should be stressed that these category labels have no role to play in the browsing framework and that no use was made of them by the algorithms.

3.2 Simulating the user

Two types of image feature were extracted: 36-dimensional colour histograms and 18-dimensional texture features based on a wavelet transformation [2]. These were concatenated to give 54-dimensional vectors.

Quantitative evaluations were performed by simulating use of the system. A fixed distance metric, $D_{\text{user}}(\boldsymbol{q}, \boldsymbol{x}; \boldsymbol{w}_{\text{user}})$ based on image features was used by a simulated user. Each simulated session was initiated by randomly selecting two images from the database, one as query and one as target. In the first iteration, the system retrieved images based on a

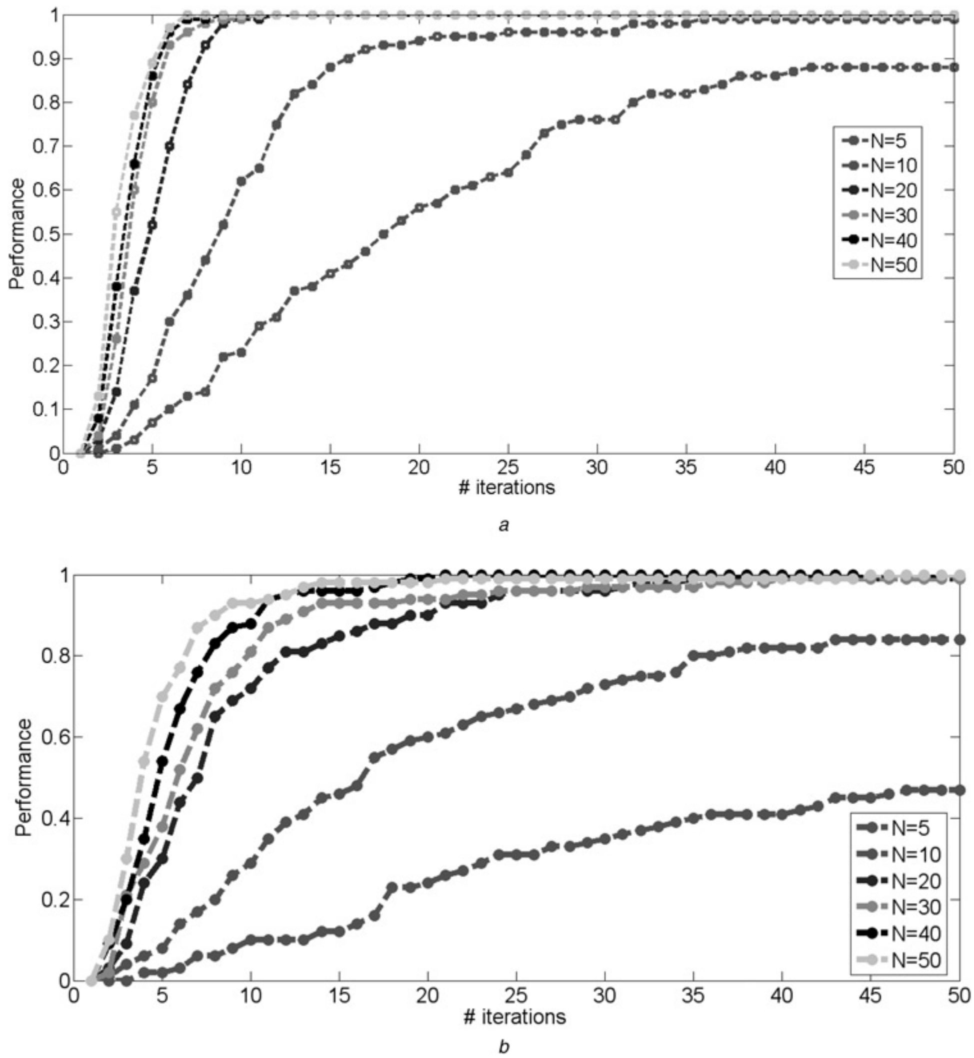

Fig. 5 *Retrieval performance*

a Performance for various valves of N
b Performance without learning; the query was reselected and $\boldsymbol{w}$ was set at random

IET Image Process., 2014, Vol. 8, Iss. 10, pp. 610–618
doi: 10.1049/iet-ipr.2013.0514

pre-specified metric $D_0(\boldsymbol{q}, \boldsymbol{x}; \boldsymbol{w}_0)$ that differed from D_{user}. At each iteration, the system retrieved N images. The simulated user then used D_{user} to select the closest retrieved image to the target as the new query. Additionally, the simulated user selected the $M-1$ next closest (most relevant) images and arranged these in terms of distance to the query. In this way, inequality constraints were generated and used by the learning algorithm to update the metric to better approximate D_{user}. The session terminated when the target was retrieved or after a maximum number of iterations. Once an image had been retrieved it was excluded from being retrieved in subsequent iterations.

The number of images retrieved, N and the number selected for ranking feedback, M, at each iteration were free parameters. Larger values of N result in a greater choice of query. Larger values of M result in more feedback information per iteration.

3.2.1 Retrieval performance when N = M: A first set of experiments investigated performance assuming that all images in each layout were deemed relevant and thus arranged by the user, that is, $N = M$. In each case this generated $N-1$ constraints at each iteration.

The distance metric was initialised to only use the texture features: weights in (1) were set to equal, non-zero values for the texture features and to zero for the colour features. In contrast, the simulated user used a distance metric in which colour features had equal, non-zero values and texture features had weights of zero.

Performance was measured as the fraction of trials in which the target was retrieved within a given number of iterations. Fig. 5*a* plots performance for different values of $N \in \{5, 10, 20, 30, 40, 50\}$. In each case, 100 sessions were simulated. Browsing was terminated after 50 iterations if still unsuccessful. Success was nearly always achieved in fewer than ten iterations for $N \geq 20$.

Fig. 5*b* shows the performance obtained without using learning but instead changing the system metric at each iteration by setting the elements of $\boldsymbol{w}$ to 0 or 1 at random. The retrieved image that was closest to the target using the new metric was selected as the query for the next iteration. The retrieval rates were inferior to those obtained using learning.

Metric learning was compared with ranking-SVM using code from SVM-Light [13]. Two naive methods were also compared: random selection of N images without any simulated user interaction, and use of the initial metric for matching throughout. Finally, the methods were compared with retrieval using the ideal metric, D_{user}. Fig. 6 shows comparative results suggesting that the metric learner was superior to ranking-SVM, especially when N was small. For example, for $N = 10$, it achieved a retrieval rate of 59% by iteration 10, whereas ranking-SVM achieved a rate of 7%. For $N = 20$, these rates increased to 96% and 50%, respectively. Retrieval rates obtained by metric learning quickly approached those obtained using the ideal metric as N increased. When $N = 40$, the rate differed notably only for those sessions that succeeded in less than five iterations, and then only by about 5%.

3.2.2 Retrieval performance when N > M: A second set of experiments investigated performance when not all images displayed in a layout were selected and arranged by the user, that is, $M < N$. The simulated user selected those M images that were most similar to the target using the current distance metric, D_t, and arranged them to yield

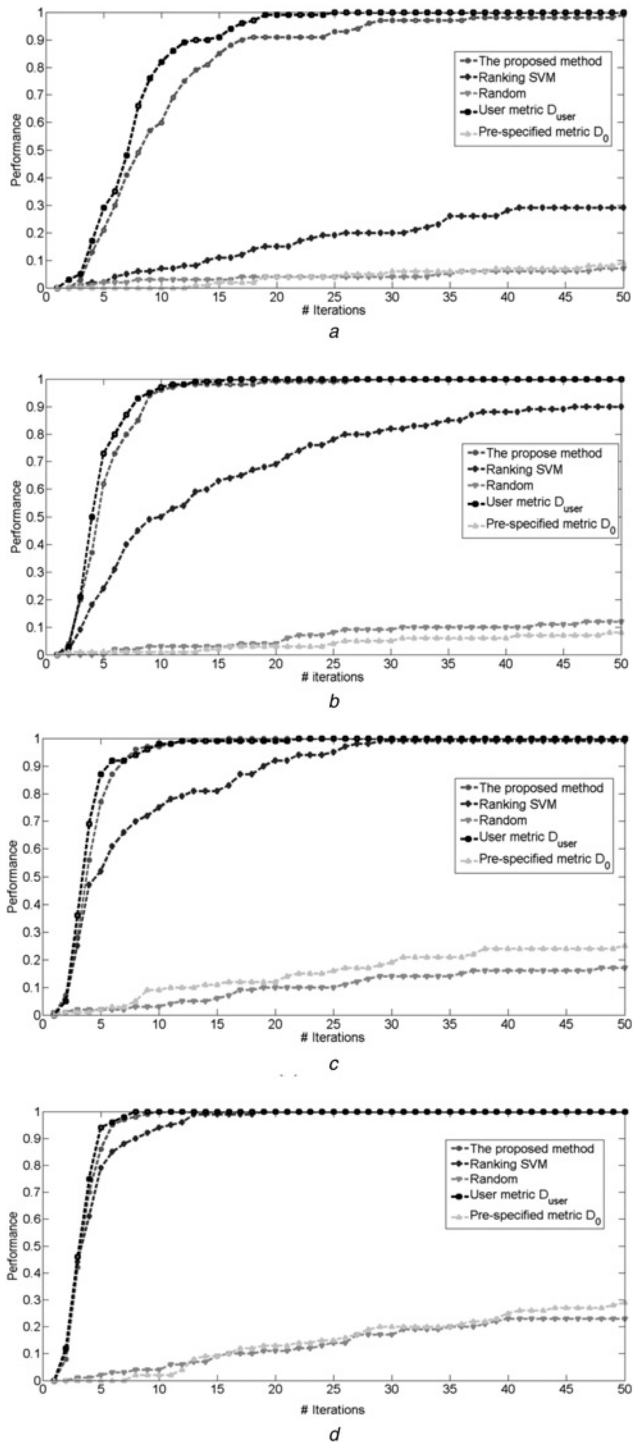

Fig. 6 *Comparison of performance of different methods for different values of N*

a N = 10
b N = 20
c N = 30
d N = 40

constraints. Fig. 7 plots performance for different values of $M \in \{3, 5, 10, 20, 30, 40\}$ when $N = 50$. Fig. 7*a* was obtained using all $N-1$ constraints in each iteration (Section 2.4). Fig. 7*b* was obtained using only the $M-1$ constraints expressing order relationships between the selected images. Comparing these two graphs, the results suggest that including relationships with non-selected images in the constraint set (i.e., using $N-1$ constraints) is helpful when the number of selected images is small (e.g., $M = 3$). These

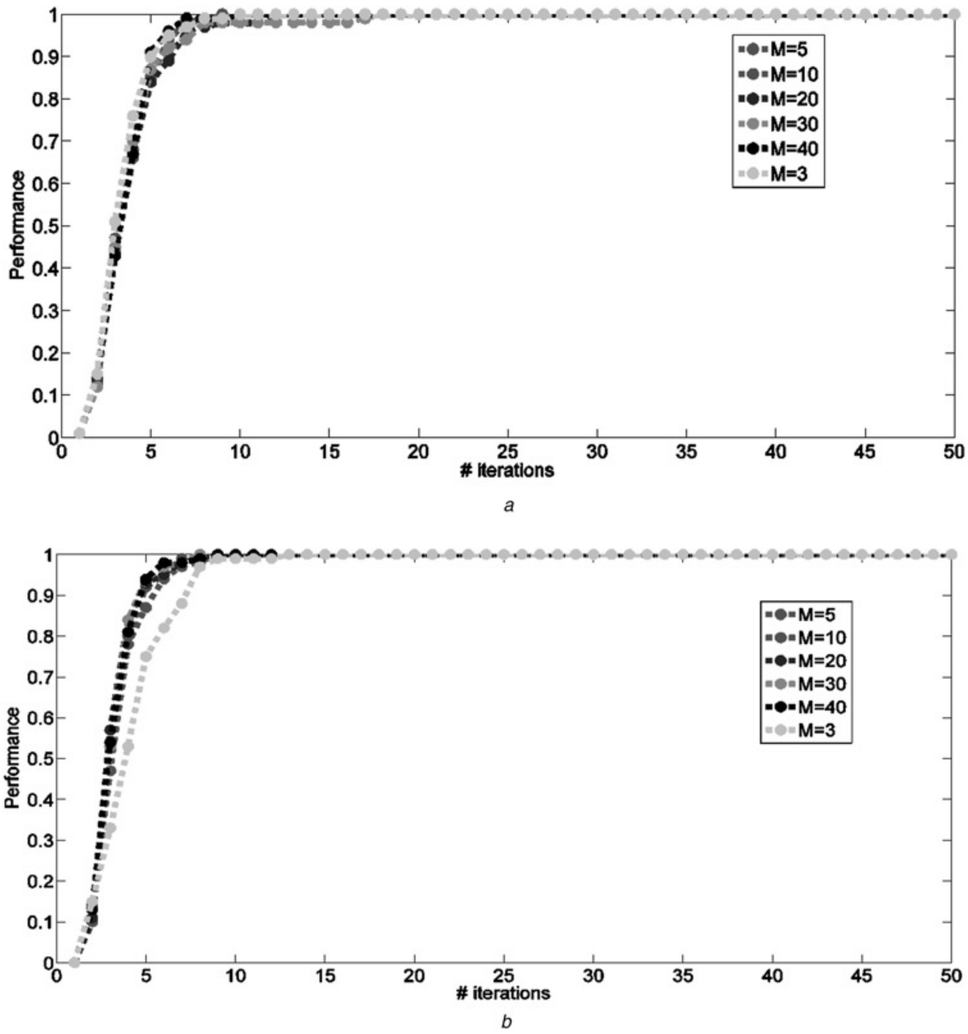

Fig. 7 *Retrieval performance with N = 50 and M < N with*
a $N-1$ constraints and
b $M-1$ constraints

results also demonstrate that the number of iterations is increased only slightly, if at all, when the user selects as few as five images for feedback compared with selecting all images.

3.2.3 Changing the query: A user might decide to choose a new query during a session because what they seek has changed, perhaps as a result of inspiration gained from browsing. However, even when what they seek has not changed, it can be useful to change the query when doing so gives a query that is closer to the 'target'. To demonstrate the benefit of doing so, the method was run without the ability to change query, that is, the query image remained fixed throughout a session. Fig. 8 shows the

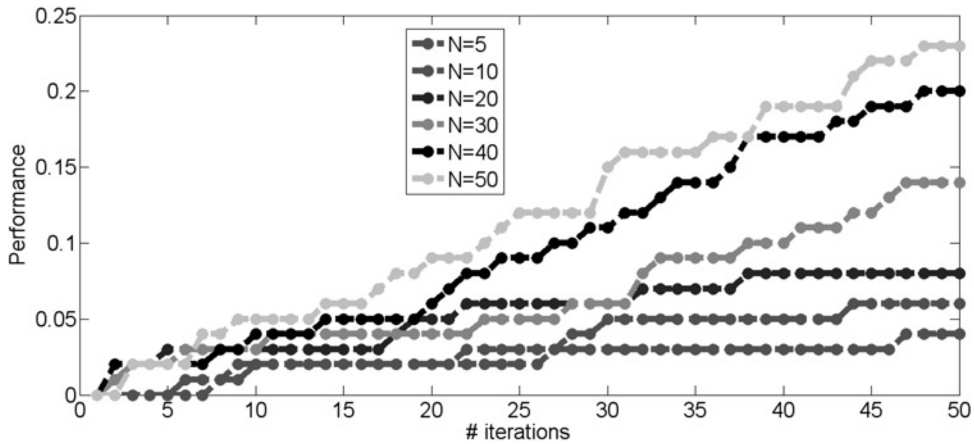

Fig. 8 *Retrieval performance with query image fixed during each session (N = M)*

result. It can be seen (cf. Fig. 5*a*) that allowing the user to change query can dramatically reduce the number of iterations needed.

3.2.4 Initial metric mismatch: We investigated the effect of mismatch between the user metric, D_{user} and the initial system metric, D_0, on retrieval performance. In this experiment, each element in the parameter vectors $\boldsymbol{w}_{\text{user}}$ and $\boldsymbol{w}_0$ was set to either 0 or 1, in effect switching features off or on. The Hamming distance between $\boldsymbol{w}_{\text{user}}$ and $\boldsymbol{w}_0$ was used as a measure of mismatch between the two metrics. Since the feature space was 54-dimensional, the Hamming distance was always in the range [0 54].

Fig. 9*a* plots the mean and standard deviation of the number of iterations per simulated browsing session when $N = M = 20$. For each value of the Hamming distance, 1000 trials were run, each with randomly generated distance metrics having that Hamming distance. No effect of mismatch between user metric and initial system metric is apparent.

3.2.5 Distance between query and target: We investigated the effect of similarity of the query image to the target image. Simulated sessions were run using randomly generated metrics (Section 3.2.4) and randomly selected query and target images. Fig. 9*b* shows a plot from 1000 trials in which the number of iterations is plotted against the distance between initial query and target as measured using D_{user}. The correlation coefficient was -0.22, suggesting that the similarity between initial query and target is not a critical factor in determining retrieval performance.

3.3 Interactive online experiment with users

Four subjects (two male, two female) tested the system. Each performed ten search sessions. Target images were selected by users and came from 36 different categories. Before each session, the system displayed a layout of 100 images selected randomly from the database. The user selected whichever of these images was most similar to the target as the initial query image unless the user did not consider any of these 100 images to be similar to the target. In the latter case, the system offered them another 100 randomly selected images from which they were forced to choose. Given the results of the simulation above, $N = M = 20$ was chosen as a reasonable trade-off. A 144-dimensional colour correlogram [1] was used to represent each image in this experiment.

Each iteration requires the user to select a query and move images to provide feedback on similarity to the query. This is more time-consuming than the CPU time for learning,

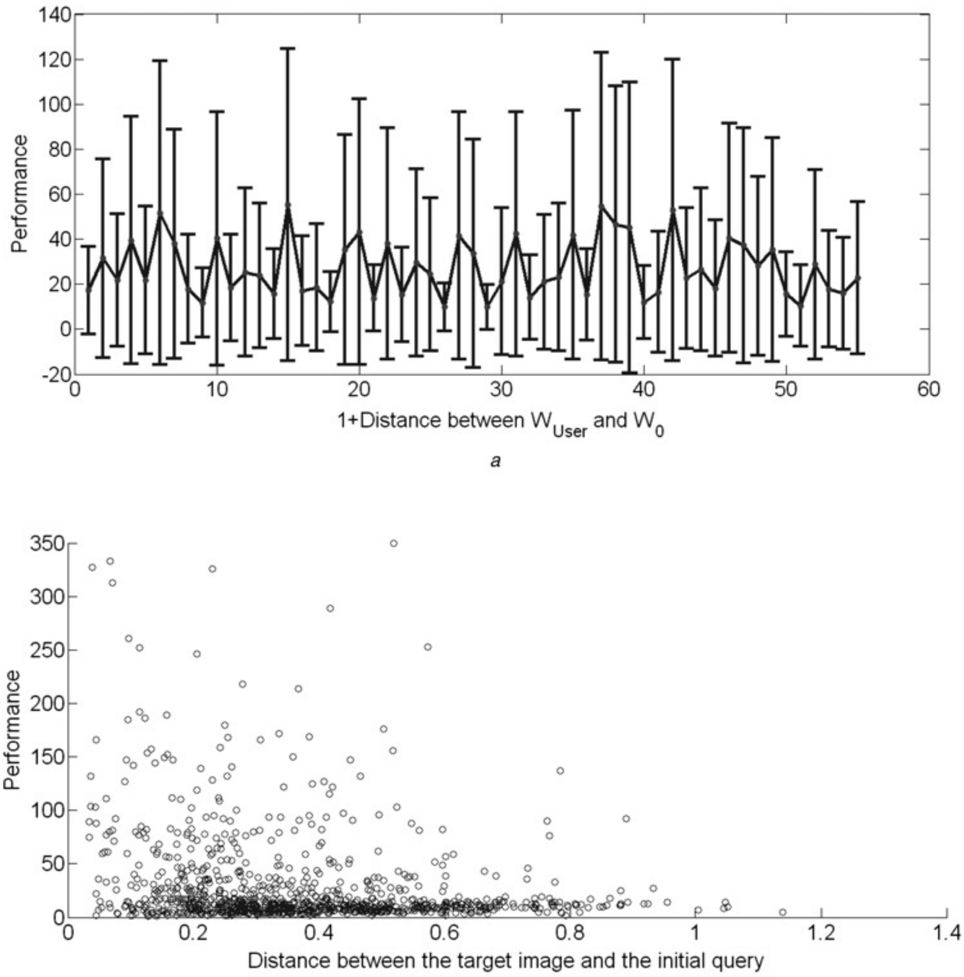

Fig. 9 *Effects of metric mismatch and query mismatch on performance*

a Performance with different Hamming distances between $\boldsymbol{w}_{\text{user}}$ and $\boldsymbol{w}_0$
b Scatter-plot of performance against distance between the initial query and the target image

matching and visualisation. Query selection normally took < 10 s whereas arranging images took 25–50 s. If a target was not found after 10 iterations, search was deemed to have failed. There were 40 search sessions in total and, of these, five failed, three found the target without any interaction other than initial query selection, 20 were successful within five iterations and 12 others were successful using more than five iterations. Overall, successful sessions required an average of five iterations to retrieve the target.

4 Conclusions and recommendations

A framework for adaptive image browsing was presented based on HELD visualisations, qualitative user feedback provided by manipulation of these, and online learning of image-image distance metrics. A method for efficient, quantitative characterisation and comparison of methods using simulation was presented. Results suggest that the approach has potential for application to real-world, interactive browsing and retrieval.

Maximal-margin metric learning based on user-provided constraints performed better than ranking-SVM in this context, and in some scenarios approached retrieval performance obtained when the user metric was known. Results suggest that such adaptation is effective even when feedback is provided through selection of only a few images (e.g., 5) per iteration. The ability to change query during a session can dramatically reduce the number of iterations. Interestingly, neither dissimilarity of the initial query to the eventual retrieval result nor mismatch of the initial system metric impacted notably on the number of iterations needed.

Although useful for efficient comparison of algorithms, evaluation using simulation has limitations. One aspect for future work is modelling user distance metrics in ways that allow them to adapt during a session. Preliminary user testing was presented here; further work is needed to fully evaluate the approach with more real users and various query scenarios.

5 Acknowledgments

We are grateful to Annette A. Ward and Ruixuan Wang for valuable discussions. This research received funding from the UK Technology Strategy Board FABRIC project, a collaboration between the University of Dundee, Liberty Art Fabrics, System Simulation and the V&A Museum. This research was done while J. Han was at University of Dundee.

6 References

1 Huang, J., Ravi Kumar, S., Mitra, M., Zhu, W., Zabih, R.: 'Spatial color indexing and applications', *Int. J. Comput. Vis.*, 1999, **35**, pp. 245–268
2 Smith, J.R., Chang, S.F.: 'Automated binary texture feature sets for image retrieval'. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Atlanta, 1996
3 Chen, J.-J., Su, C.-R., Grimson, W.E.L., Liu, J.-L., Shiue, D.-H.: 'Object segmentation of database images by dual multiscale morphological reconstructions and retrieval applications', *IEEE Trans. Image Process.*, 2012, **21**, (2), pp. 828–843
4 Tong, S., Chang, E.: 'Support vector machine active learning for image retrieval'. ACM Conf. on Multimedia, 2001
5 He, X., Ma, W.Y., Zhang, H.J.: 'Learning an image manifold for retrieval'. ACM Int. Conf. on Multimedia, New York, 2004
6 Zhang, L., Wang, L., Lin, W.: 'Semi-supervised biased maximum margin analysis for interactive image retrieval', *IEEE Trans. Image Process.*, 2012, **21**, (4), pp. 2294–2308
7 Wang, B., Pan, F., Hu, K., Paul, J.: 'Manifold-ranking based retrieval using k-regular nearest neighbor graph', *Pattern Recognit.*, 2012, **45**, (4), pp. 1569–1577
8 Han, J., Ngan, K., Li, M., Zhang, H.: 'A memory learning framework for effective image retrieval', *IEEE Trans. Image Process.*, 2005, **14**, (4), pp. 511–524
9 Si, L., Jin, R., Hoi, S., Lyu, M.: 'Collaborative image retrieval via regularized metric learning', *Multimedia Syst.*, 2006, **12**, (1), pp. 34–44
10 Yang, Y., Nie, F., Xu, D., Luo, J., Zhuang, Y., Pan, Y.: 'A multimedia retrieval framework based on semi-supervised ranking and relevance feedback', *IEEE Trans. Pattern Anal. Mach. Int.*, 2012, **34**, (4), pp. 723–742
11 Zhang, L., Wang, L., Lin, W.: 'Conjunctive patches subspace learning with side information for collaborative image retrieval', *IEEE Trans. Image Process.*, 2012, **21**, (8), pp. 3707–3720
12 Cox, I.J., Miller, M.L., Minka, T.P., Papathomas, T.V., Yianilos, P.N.: 'The Bayesian image retrieval system, Pichunter: Theory, implementation, and psychophysical experiments', *IEEE Trans. Image Process.*, 2000, **9**, (1), pp. 20–37
13 Joachims, T.: 'Optimizing search engines using clickthrough data'. ACM SIGKDD Int. Conf. on Knowledge Discovery and DataMining, Alberta, Canada, 2002
14 Schultz, M., Joachims, T.: 'Learning a distance metric from relative comparisons'. Neural Information Processing Systems, Berlin, 2003
15 Freund, A., Iyer, R., Schapire, R.E., Lozano-Perez, T.: 'An efficient boosting algorithm for combining preferences', *J. Mach. Learn. Res.*, 2003, **4**, pp. 933–969
16 Frome, A.: 'Learning Distance Functions for Examplar-based Object Recognition', Ph.D. thesis. UC Berkeley, 2007
17 Hu, Y., Li, M., Yu, N.: 'Multiple-instance ranking: Learning to rank images for image retrieval'. IEEE Conf. on Computer Vision and Pattern Recognition, Anchorage, USA, 2008
18 Lee, J.E., Jin, R., Jain, A.K.: 'Rank-based distance metric learning: An application to image retrieval'. IEEE Conf. on Computer Vision and Pattern Recognition, Anchorage, 2008
19 Faria, F.F., Veloso, A., Almeida, H.M., Valle, E., Torres, R.d.S., Goncalves, M.A., Meira, W.Jr.: 'Learning to rank for content-based image retrieval'. Int. Conf. on Multimedia Information Retrieval, 2010
20 Huang, W., Chan, K.L., Li, H., Lim, J.H., Liu, J., Wong, T.Y.: 'Content-based medical image retrieval with metric learning via rank correlation'. Int. Workshop on Machine Learning in Medical Imaging, 2010, pp. 18–25
21 Wang, G., Forsyth, D., Hoiem, D.: 'Comparative object similarity for improved recognition with few or no examples'. IEEE Conf. on Computer Vision and Pattern Recognition, 2010
22 Parikh, D., Grauman, K.: 'Relative Attributes'. IEEE Int. Conf. on Computer Vision, 2011
23 Rodden, K.: 'Evaluating Similarity-Based Visualisations as Interfaces for Image Browsing', Ph.D. Thesis. University of Cambridge, 2001
24 Moghaddam, B., Tian, Q., Lesh, N., Shen, C., Huang, T.S.: 'Visualization and user-modeling for browsing personal photo libraries', *Int. J. Comput. Vis.*, 2004, **56**, pp. 109–130
25 Wang, R., McKenna, S.J., Han, J., Ward, A.A.: 'Visualizing image collections using high-entropy layout distributions', *IEEE Trans. Multimedia*, 2010, **12**, (8), pp. 803–813
26 Han, J., McKenna, S.J., Wang, R.: 'Learning query-dependent distance metrics for interactive image retrieval'. Int. Conf. on Computer Vision Systems, Liege, 2007
27 Arya, S., Mount, D.M., Netanyahu, N.S., Silverman, R., Wu, A.Y.: 'An optimal algorithm for approximate nearest neighbor searching', *J. ACM*, 1998, **45**, pp. 891–923

doi: 10.1049/iet-ipr.2013.0514