2009

Content based image retrieval using unclean positive examples

Jun Zhang
University of Wollongong

Lei Ye
University of Wollongong, lei@uow.edu.au

Content based image retrieval using unclean positive examples

Abstract

Conventional content-based image retrieval (CBIR) schemes employing relevance feedback may suffer from some problems in the practical applications. First, most ordinary users would like to complete their search in a single interaction especially on the Web. Second, it is time consuming and difficult to label a lot of negative examples with sufficient variety. Third, ordinary users may introduce some noisy examples into the query. This correspondence explores solutions to a new issue that image retrieval using unclean positive examples. In the proposed scheme, multiple feature distances are combined to obtain image similarity using classification technology. To handle the noisy positive examples, a new two-step strategy is proposed by incorporating the methods of data cleaning and noise tolerant classifier. The extensive experiments carried out on two different real image collections validate the effectiveness of the proposed scheme.

Disciplines

Physical Sciences and Mathematics

Publication Details

Zhang, J. & Ye, L. (2009). Content based image retrieval using unclean positive examples. IEEE Transactions on Image Processing, 18 (10), 2370-2375.

Content Based Image Retrieval Using Unclean Positive Examples

Jun Zhang and Lei Ye, *Senior Member, IEEE*

***Abstract*—Conventional content-based image retrieval (CBIR) schemes employing relevance feedback may suffer from some problems in the practical applications. First, most ordinary users would like to complete their search in a single interaction especially on the web. Second, it is time consuming and difficult to label a lot of negative examples with sufficient variety. Third, ordinary users may introduce some noisy examples into the query. This correspondence explores solutions to a new issue that image retrieval using unclean positive examples. In the proposed scheme, multiple feature distances are combined to obtain image similarity using classification technology. To handle the noisy positive examples, a new two-step strategy is proposed by incorporating the methods of data cleaning and noise tolerant classifier. The extensive experiments carried out on two different real image collections validate the effectiveness of the proposed scheme.**

***Index Terms*—Classifier combination, content-based image retrieval (CBIR), feature aggregation, noise tolerant, support vector machine (SVM).**

I. Introduction

Content-based image retrieval (CBIR) is a technique to search for images relevant to the user's query from an image collection [1]. In the last decade, the conventional CBIR schemes employing relevance feedback have achieved certain success [2]. The idea of relevance feedback is to involve the user in the retrieval process so as to improve the final retrieval results. Normally, the user labels some returned images as relevant or irrelevant and the system adjusts the retrieval parameters based on the user's feedback. Relevance feedback can go through one or more iterations until the user is satisfied with the results. However, the conventional CBIR schemes employing relevance feedback may suffer from some problems in practical applications. First, if not impossible, ordinary users have little patience to persist in the feedback iterations, and most would like to complete their search in a single interaction [3]. Second, labeling some positive (relevant) examples is easy while labeling sufficient negative (irrelevant) examples is time consuming and difficult [4]. Third, some noisy examples may present since ordinary users normally have no expertise in constructing a high quality query. To the best of our knowledge, most existing retrieval schemes fail to address the problem of noisy examples. In this correspondence, we explore solutions to a new issue that image retrieval using unclean positive examples. The user supplies several unclean positive examples as a query and the CBIR system will return the relevant images from an image collection in a single interaction. Under this circumstance, some noisy positive examples may present in the query which are irrelevant images mislabeled by the user [5] or weakly relevant images which can not well represent the set of relevant images. The noisy examples will affect the image retrieval performance seriously. The solution of this problem is useful for the practical applications of CBIR in which relevance feedback is not a suitable choice but some unclean positive examples can be provided.

In the proposed scheme, the similarity between two images is obtained by combining the distances on multiple visual features, such as, color and texture [6], named feature aggregation. We propose a new way to perform feature aggregation, instead of existing heuristic methods [7]–[11]. Let a query image as the prototype, a new feature dissimilarity space is constructed in which an image is represented using the feature distances to the prototype. Then, feature aggregation can be formulated as a binary classification problem and solved by conventional classification technologies. To handle the noisy positive examples, a new two-step strategy is proposed by incorporating the methods of data cleaning and noise tolerant classifier. In step 1, an ensemble of support vector machines (SVMs) [12], [13] are trained in a feature dissimilarity space corresponding to a reliable positive example, which are used as consensus filters to identify and eliminate the noisy positive examples. To train SVMs, some negative training examples are randomly labeled from the image collection. In step 2, each retained positive example is associated with a relevance probability to further alleviate the influence of the retained noisy positive examples. The similarities of an image to the retained positive examples are then combined to get the final image relevance for ranking.

The remainder of correspondence is organized as follows. Some related work is briefly reviewed in Section II. Section III presents a novel feature aggregation method. Section IV proposes a two-step strategy to handle noisy positive examples. A large number of experiments are reported in Section V. Conclusions are drawn in Section VI.

II. Related Work

A. Feature Aggregation

Feature aggregation is an approach to get image similarity by combining multiple feature distances. In the literatures, various fixed aggregation functions have been applied and evaluated in feature aggregation methods [7]–[9]. The experiments show a proper aggregation function is much important to retrieval performance. MARS [10] represented the user query as a boolean expression over visual features, and similarity between images becomes the evaluation of this expression using feature distances. To extend the traditional Boolean model, Kushki *et al.* [11] proposed a hierarchical decision fusion framework using fuzzy logic to combine multiple feature distances. The problem of these methods is requiring the system designer or ordinary users to manually tune the internal parameters. Rui *et al.* [14] presented an optimization formulation to learn the users' preference, which computed feature weighting automatically using multiple positive examples, but it focuses on linear aggregation function and does not use the information of negative examples.

B. Learning From Positive and Unlabeled Examples

Recently, learning from positive and unlabeled examples has got much attention in text classification. The key feature of this problem is that there are no labeled negative examples, which makes conventional supervised or semi-supervised learning techniques inapplicable. One popular approach takes a two-step strategy. In step 1, a set of reliable negative examples are identified from the unlabeled set. The existing methods include the naive Bayesian technique (NB) used in [15], the Rocchio technique used in Roc-SVM [16], the Spy technique used in S-EM [17] and 1-DNF technique used in PEBL [18]. In step 2, a set of classifiers are built by iteratively applying a classification algorithm and then selecting a good classifier from the set. The existing

Manuscript received December 18, 2008; revised May 27, 2009. First published July 06, 2009; current version published September 10, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Dan Schonfeld.

The authors are with the School of Computer Science and Software Engineering, University of Wollongong, Wollongong, NSW, 2522 Australia (e-mail: jz484@uow.edu.au; lei@uow.edu.au).

Digital Object Identifier 10.1109/TIP.2009.2026669

methods include direct SVM used in [15], EM used in S-EM, SVM iteratively used in PEBL and SVM iteratively and selecting a final classifier used in Roc-SVM. These research show it is reasonable to get negative examples from unlabeled examples. However, this approach based on critical analysis is time consuming and not suitable to the real-time applications of CBIR.

C. Classification Using Unclean Training Examples

In machine learning, classification using unclean training examples is an open issue [19], [20]. There are two main approaches to address this issue, data cleaning and noise tolerant classifier. Since bad examples can be removed prior to classifier induction, data cleaning may increase the classification accuracy. The boosting algorithm [21] can be used to avoid the noise influence on constructing the classifier via combining a set of classifiers. In [20], an ensemble method based approach was proposed to identify and eliminate mislabeled training examples for supervised learning. The analytical and empirical evaluation shows that consensus filters are conservative at throwing away good data at the expense of retaining bad data and that majority filters are better at detecting bad data at the expense of throwing away good data. So consensus filters are suitable to a paucity of data and majority vote filters are preferable for an abundance of data. In the other approach, some efforts have been taken to construct noise tolerant classifiers directly, which have no potential risk of removing good examples. In [22], a noise generative model was introduced into kernel fisher discriminant analysis to handle noisy examples. The key idea was to alleviate the noise influence by associating with each example a probability of the label being flipped.

III. Novel Feature Aggregation Method

In this section, we present a new classification-based feature aggregation method.

A. Feature Dissimilarity Space

Let us consider an image collection $\mathcal{I}$ containing N images, $\mathcal{I} = \{I_1, I_2, \ldots, I_N\}$. Assuming m visual features are designed, the feature representation of an image I is a set of m feature vectors, $\{F_{\cdot k}\}_{k=1}^{m}$, in high-dimensional feature spaces. A user supplies some positive example images as a query, $\mathcal{P} = \{P_i\}_{i=1}^{p}$.

Let a query image $P \in \mathcal{P}$ as the prototype, we construct a new feature dissimilarity space by modifying the method proposed by Duin and Pekalska [23], [24]. For each collection image I_i, we have

$$S_i = (s_{i1}, s_{i2}, \ldots, s_{im}) \tag{1}$$

where s_{ij} represents the dissimilarity between I_i and P on the jth feature and S_i is a vector in an m-dimensional space $\mathcal{S}$, called feature dissimilarity space. In this correspondence, the dissimilarity is defined by a feature distance. We denote $D_j(\cdot,\cdot)$ as a specified distance metric for the jth visual feature, then

$$s_{ij} = D_j(F_{ij}, F_{Pj}). \tag{2}$$

Therefore, all images in $\mathcal{I}$ are vectors in $\mathcal{S}$.

There are some differences between feature dissimilarity space and conventional dissimilarity space [23]–[26]. First, feature dissimilarity space is introduced to address the feature aggregation problem which has only one prototype, while conventional dissimilarity space has multiple prototypes selected by the system designer. Second, a collection image is represented using multiple feature distances to the prototype in feature dissimilarity space, while in conventional dissimilarity space a point is represented using the distances to multiple prototypes. Third, the dimension of a feature dissimilarity space depends on the number of visual features applied in a CBIR system, while in a conventional dissimilarity space, the dimension depends on the number of prototypes.

Compared with original feature space, feature dissimilarity space inherits the advantages of conventional dissimilarity space. Sometimes it is difficult to create a combined feature space with a unified distance metric for multiple features, but we always can create a feature dissimilarity space [11], [24]. In feature dissimilarity space, feature aggregation can be transformed into a classification problem and addressed by conventional classification technologies. Such that image retrieval using feature aggregation can be optimized.

B. SVM-Based Feature Aggregation

In this correspondence, we cast feature aggregation as a binary classification problem. The positive class consists of relevant images to the query and the negative class consists of all irrelevant images. SVM algorithm [12], [13] is chosen to design the binary classifier because of its good generalization and noise tolerant ability.

Consider a linear separable binary classification problem in feature dissimilarity space with n training examples, $\{(S_i, y_i)\}_{i=1}^{n}$ and $y_i = \{+1, -1\}$, where S_i is a training example and y_i is the label of this example. The query images are labeled by $+1$ and some images in the collection are randomly labeled by -1. SVM separates the positive class and negative class by a hyperplane, $W \cdot S + b = 0$, where S is an input vector, W and b are the hyperplane coefficients and scalar. The goal in training an SVM is to find the separating hyperplane with the largest margin, which is represented as

$$\begin{cases} y_i(W \cdot S_i + b) \geq +1, i \in [1, n] \\ \min \|W\|^2/2. \end{cases} \tag{3}$$

The solution can be found through a Wolfe dual problem with the undetermined Lagrangian multipliers α_i. To get a potentially better representation of the data, the data points can be mapped into a higher dimensional space using the proper chosen nonlinear ϕ-functions, $K(S_i, S_j) = \phi(S_i) \cdot \phi(S_j)$, where $K(\cdot)$ is a kernel function. Then, we get the kernel version of the Wolfe dual problem. Thus, for a given kernel function, the output hyperplane decision function of SVM is

$$f(S) = \sum_{i=1}^{n} \alpha_i y_i K(S_i, S) + b. \tag{4}$$

The SVM classifier is given by, $C(S) = \mathrm{sgn}(f(S))$. To deal with cases where there may be no separating hyperplane, the soft margin SVM can be applied, the goal of which can be expressed as

$$\begin{cases} y_i(W \cdot S_i + b) \geq 1 - \xi_i, \xi_i \geq 0,\ i \in [1, n] \\ \min \|W\|^2/2 + C \sum_{i=1}^{n} \xi_i \end{cases} \tag{5}$$

where ξ_is are slack variables, $\sum_{i=1}^{n} \xi_i$ is an upper bound on the number of training errors and $C \geq 0$ is a parameter to control the penalty to errors.

In this method, the output of SVM, $f(x)$, is used as the result of feature aggregation, i.e., image similarity. Based on the kernel trick, a linear feature aggregation method can be easily extended to a nonlinear one.

IV. Proposed Image Retrieval Scheme

In this section, we propose a new image retrieval scheme to handle noisy positive examples, which consists of two steps, noise identification and elimination, and noise tolerant relevance calculation.

A. Noise Identification and Elimination

First, an ensemble of SVMs as consensus filters [20] is constructed to filter out the noisy positive examples. Our goal is to remove some bad examples as well as retain all good examples. To create a feature dissimilarity space, we need to select a prototype. Instead of random selection and average point strategies which are sensitive to noise, we apply a new strategy to choose a reliable positive image as the prototype based on the idea of k-medoids approach [27]. The strategy is more robust to noisy examples, which can be represented as

$$P_o = \arg\min_{P \in \mathcal{P}} \sum_{i=1}^{n_p} \sum_{j=1}^{m} s_{ij}^{(p)} \tag{6}$$

where $s_{ij}^{(p)}$ is the distance of the ith positive examples P_i to a prototype candidate P on the jth feature.

Since only positive examples are available, the conventional supervised or semi-supervised learning techniques are inapplicable. One popular approach is to label some reliable negative examples from the unlabeled data through critical analysis [15]–[18]. However, this approach is time consuming and not suitable for real-time image retrieval. In our scheme, random sampling is applied to label some images in the collection as negative examples. In a large image collection, it has a high correct probability to label a random image as a negative example. Furthermore, a large set of random negative examples can benefit the retrieval performance [4].

In the feature dissimilarity space, the positive examples and an equal number of negative examples can be used to train support vector machine (SVM). Since the noisy examples present, an SVM classifier trained in feature dissimilarity space will be unstable. We use different negative example sets to train SVM and get multiple classifiers. A similar strategy, named asymmetric bagging, has been applied in [28], which can effectively handle the unstable and unbalance classifiers. After that, we apply all classifiers to classify the positive examples provided by a user. Based on the consensus strategy, the examples labeled by all SVM classifiers as negative will be identified as noisy positive examples and eliminated.

B. Noise Tolerant Relevance Calculation

To further handle the retained noisy positive examples after consensus filtering, we propose a noise tolerant relevance calculation method, which estimates a relevance probability for each retained positive example [22].

1) Relevance Probability Estimation: To estimate the relevance probability of an image, we propose an ensemble-based estimation algorithm, which can be regarded as a by-produce of consensus filtering. First, the sigmoid function combined with the output of an SVM classifier is used to estimate the class-conditional probability [29] for a positive example S by

$$P(L_k|C,S) = \frac{1}{1+\exp\left(-|f(S)|\right)} \tag{7}$$

where $f(S)$ is the decision value and L_k is the predicted class label, both of them are produced by an SVM classifier. In a binary classification task, L_0 and L_1 denote positive and negative class, respectively. We apply the SVM classifiers trained in Section IV-A to classify the retrained positive examples. Then, all outputs are then combined to get the conditional probabilities based on Bayes sum rule (BSR)

$$P(R|S) = \frac{1}{T} \sum_{i=1}^{T} P(L_0|C_i,S) \tag{8}$$

where $P(R|S)$ is the estimated relevance probability and T is the number of SVMs.

2) Image Relevance Calculation: In this correspondence, the similarity of an image to a query image is represented by an ensemble of SVMs. We combine multiple ensembles of SVMs to obtain the image relevance to a user's query. The relevance calculation algorithm is summarized in Table I. Three aggregation models [30] are evaluated in this correspondence.

TABLE I
RELEVANCE CALCULATION

Input: retained query image set $\mathcal{P}'$, SVM classifier V,
integer T (the number of bagging classifiers),
and the image collection $\mathcal{I}$
1. For $i = 1$ to $|\mathcal{P}'|$ {
2. $\mathcal{S}^i$ = create feature dissimilarity space for $P_i \in \mathcal{P}'$
3. For $j = 1$ to T {
4. $\mathcal{N}_{ij}$ = random example from $\mathcal{I}$, with $|\mathcal{N}_{ij}| = |\mathcal{P}'|$
5. $C_{ij} = V(\mathcal{P}', \mathcal{N}_{ij})$
6. }
7. }
6. $C^*(\mathcal{I})$ = classifier combination $\{C_{ij}(\mathcal{I})\}$
Output: C^*

- *SVM-Weighted-MVR*: For a given image, we first use the weighted majority vote rule (MVR) to recognize it as query relevant or irrelevant. The weighted MVR can be represented as

$$C^*(I) = \operatorname{sgn}\left(\sum_{i,j} w_i \cdot C_{ij}(I) - \frac{T\sum_i w_i}{2}\right) \tag{9}$$

  where I is a collection image, $C_{ij}(I)$ is a sign, 0 or 1, produced by the jth classifier for the ith retained positive example, and w_i is the weighting assigned to this example. In this correspondence, $w_i = P(R|S)$ represents the relevance of a positive example. Then, we measure the relevance between the image and the query as the output of the individual SVM classifier, which gives the same label as the weighted MVR and produces the highest weighted confidence value (the weighted absolute value of the decision function of the SVM classifier).
- *SVM-Weighted-BSR*: For a given image, we first use the weighted BSR to recognize it as query relevant or irrelevant. The weighted BSR can be represented as

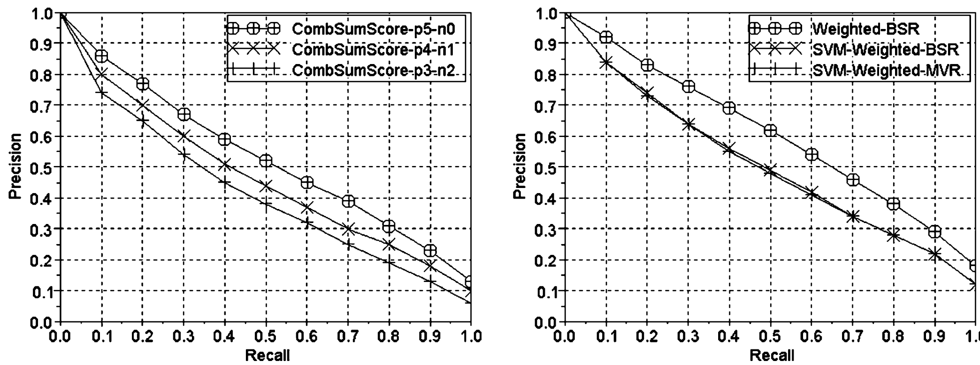
$$C^*(I) = \arg\max_k \sum_{i,j} w_i \cdot P(L_k|C_{ij}, I) \tag{10}$$

  where $P(L_k|C_{ij}, I)$ represents the class-conditional probability which can be computed by (7). Then, we measure the relevance between the image and the query using the individual SVM classifier, which gives the same label as the weighted BSR and has the highest weighted confidence value.
- *Weighted-BSR*: The output of the weighted BSR, $\sum_{i,j} w_i \cdot P(L_k|C_{ij}, I)$, can be directly used as a relevance measure between a given image and the query.

The aggregation models without weighting have been evaluated and reported in [28]. That work does not consider the noisy examples, so $w_i = 1$. In this correspondence, the weighting is used to alleviate the noise influence.

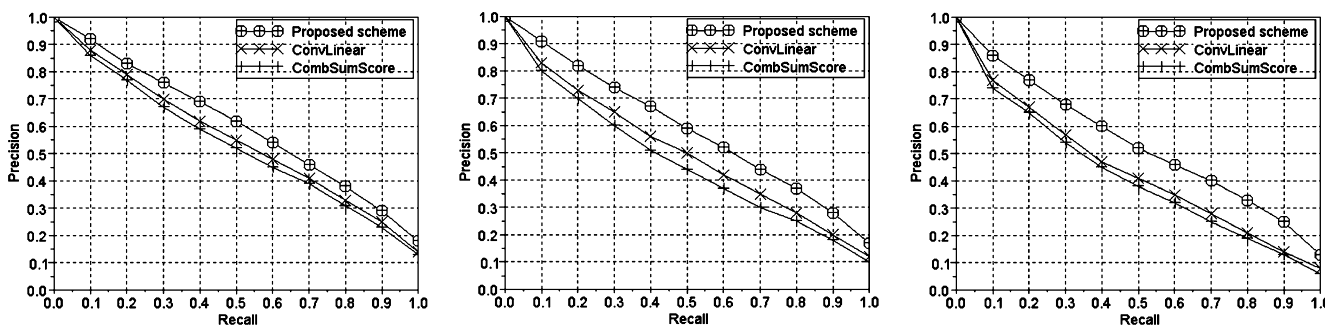
V. EXPERIMENTAL EVALUATION

In the experimental evaluation, a query consists of several positive examples and retrieval results are returned in a single interaction. Two state-of-the-art feature aggregation based retrieval schemes are implemented for comparison, CombSumScore and ConvLinear. CombSumScore is the best one in all schemes evaluated by Donald *et al.* [9] for multiple features and multiple examples. In CombSumScore, the similarity between two images are represented using the average of multiple normalized feature distances, and the image relevance to the query is



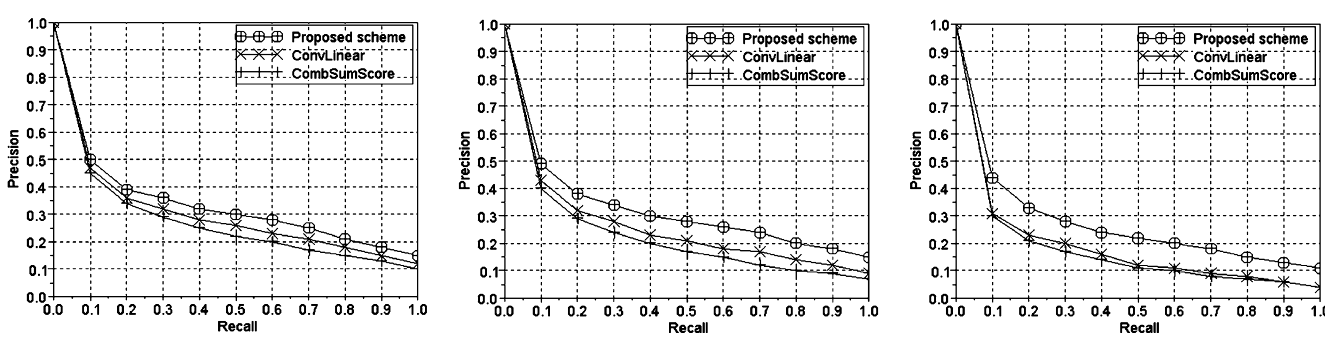
(a) (b)

Fig. 1. Evaluation of noise influence and aggregation models. (a) Noise influence; (b) aggregation models.



(a) (b) (c)

Fig. 2. Retrieval performance on Corel image collection. (a) No mislabeled positive example; (b) one mislabeled positive example; (c) two mislabeled positive examples.



(a) (b) (c)

Fig. 3. Retrieval performance on IAPR image collection. (a) No mislabeled positive example; (b) one mislabeled positive example; (c) two mislabeled positive examples.

defined as the average of multiple similarities to the query images. ConvLinear enhances CombSumScore by applying the linear weighting method [14] to combine multiple feature distances. In the experiments, five standardized MPEG-7 visual descriptors [6], [11] are selected for image representation. The recommended distance metrics are also used to measure the feature distances. For practical applications, each query includes five positive example images. It should be pointed out that the mechanism of handling noisy positive examples will not affect the speed of image retrieval. Since we have only a few query images to deal with, the computation time of the proposed algorithms is not high.

A. Experimental Results With Corel Images

All experiments are carried out on two different real-world image collections. The first one consists of 20 image categories and each category includes 100 Corel images. The images in a category have the same perceptual meaning, so the ground truth is based on the image category. The retrieval performances in terms of average precision and recall [31] on 300 randomly created queries are reported. We use the SVM-Light [32] to solve SVMs. Gauss kernel and default parameters are applied in our experiments. It is well known that the parameter tuning is important to SVM-based methods with nonlinear kernel. However, the practical parameter tuning methods, such as, grid search, are really time consuming. Considering the real time image retrieval, we choose to not tune parameters in the proposed scheme. Since it was reported that the number of bagging classifiers does not affect the retrieval performance [28] and our experiments also confirm this fact, ten SVMs are chosen based on experimental results.

1) Evaluation of Noise Influence: To highlight the influence of noisy positive examples to the retrieval performance, we manually introduce some mislabeled examples into the query. CombSumScore scheme is chosen for this experiment since it is a simple and effective one. Fig. 1(a) shows the precision and recall curves when different number of mislabeled examples are present. For instance, CombSumScore-p4-n1 means there are 4 true positive examples and 1 mislabeled positive example. The results demonstrate that the noisy examples can affect the retrieval performance dramatically.

2) Evaluation of Aggregation Models: In this experiment, no mislabeled positive examples are introduced. The retrieval performances of the proposed scheme with different aggregation models are reported in Fig. 1(b). From the figure, we see that Weighted-BSR can outperform SVM-Weighted-BSR and SVM-Weighted-MVR significantly. The reason may be that SVM-Weighted-BSR and SVM-Weighted-MVR choose the best individual SVM to measure the relevance. In the case of small examples, any SVM is too weak to be able to measure the relevance individually. While Weighted-BSR can aggregate the outputs of all weak SVMs to get a more confidently decision score for relevance measurement.

3) Evaluation of Image Retrieval Schemes: This experiment evaluates the retrieval schemes using unclean positive examples. The Weighted-BSR aggregation model is chosen for our scheme in accordance with the previous experimental results. The results in Fig. 2 show that the proposed scheme outperforms ConvLinear and CombSumScore especially when the number of mislabeled positive examples increases. The reason is the proposed scheme can handle noisy examples while other schemes can't. When the noisy positive examples present in a query, ConvLinear is hardly to improve the retrieval performance, since the method to compute feature weighting will fail.

B. Experimental Results With IAPR TC-12 Images

To further evaluate the retrieval performance of the proposed scheme, we performed a large number of experiments on the IAPR TC-12 benchmark image collection (ImageCLEF2006) [33] which contains 20,000 photographic images. Based on the queries and their ground truth sets defined in the CLEF Cross-language Image Track 2006, we build up 40 ground truth sets for our experiments. 500 queries are selected randomly from the defined ground truth and each query consists of five positive example images. The Weighted-BSR aggregation model is chosen for our scheme. Average precision and recall curves are reported in Fig. 3. The experimental results confirm the effectiveness of the proposed scheme.

VI. Conclusions

We addressed a new issue that image retrieval using unclean positive examples. In the proposed scheme, feature aggregation was formulated as a binary classification problem and solved by support vector machine (SVM) in a feature dissimilarity space. Incorporating the methods of data cleaning and noise tolerant classifier, a new two-step strategy was proposed to handle the noisy positive examples. In step 1, an ensemble of SVMs trained in a feature dissimilarity space are used as consensus filters to identify and eliminate the noisy positive examples. In step 2, the noise tolerant relevance calculation was performed, which associated each retained positive example with a relevance probability to further alleviate the noise influence. A large number of experiments were carried out on a sub-set of Corel image collection and the IAPR TC-12 benchmark image collection. The experimental results show that the proposed scheme outperforms the competing feature aggregation based image retrieval schemes when noisy positive examples present in the query.

References

[1] 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.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.

[2] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: A power tool for interactive content-based image retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 644–655, Sep. 1998.

[3] C. D. Manning, P. Raghavan, and H. Schutze, *Introduction to Information Retrieval*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[4] K. Tieu and P. Viola, "Boosting image retrieval," *Int. J. Comput. Vis.*, vol. 56, no. 1/2, pp. 17–36, 2004.

[5] X. He, O. King, W.-Y. Ma, M. Li, and H.-J. Zhang, "Learning a semantic space from user's relevance feedback for image retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 1, pp. 39–48, Jan. 2003.

[6] B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Jun. 2001.

[7] J. R. Smith, S. Srinivasan, A. Amir, S. Basu, G. Iyengar, C.-Y. Lin, M. Naphade, D. Ponceleon, and B. Tseng, "Integrating features, models, and semantics for trec video retrieval," in *Proc. Text Retrieval Conf.*, 2001, pp. 240–249.

[8] B. L. Tseiig, C.-Y. Lin, M. Naphade, A. Natsev, and J. R. Smith, "Normalized classifier fusion for semantic visual concept detection," in *Proc. IEEE Int. Conf. Image Processing*, Sep. 2003, vol. 2, pp. 535–538.

[9] K. M. Donald and A. F. Smeaton, "A comparison of score, rank and probability-based fusion methods for video shot retrieval," in *Proc. IEEE Int. Conf. Image And Video Retrieval*, Jul. 2005, pp. 61–70.

[10] M. Ortega, Y. Rui, K. Chakrabarti, K. Porkaew, S. Mehrotra, and T. Huang, "Supporting ranked boolean similarity queries in mars," *IEEE Trans. Knowl. Data Eng.*, vol. 10, no. 6, pp. 909–925, Jun. 1998.

[11] A. Kushki, P. Androutsos, K. N. Plataniotis, and A. N. Venetsanopoulos, "Retrieval of images from artistic repositories using a decision fusion framework," *IEEE Trans. Image Process.*, vol. 13, no. 3, pp. 277–292, Mar. 2004.

[12] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.

[13] J. C. Burges, "A tutorial on support vector machine for pattern recognition," *Data Mining Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, 1998.

[14] Y. Rui and T. S. Huang, "Optimizing learning in image retrieval," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Jun. 2000, vol. 1, pp. 1236–1243.

[15] B. Liu, Y. Dai, X. Li, W. Lee, and P. Yu, "Building text classifiers using positive and unlabeled examples," in *Proc. IEEE Int. Conf. Data Mining*, Nov. 2003, pp. 179–186.

[16] X. Li and B. Liu, "Learning to classify text using positive and unlabeled data," in *Proc. Int. Joint Conf. Artificial Intelligence*, Aug. 2003, pp. 587–592.

[17] B. Liu, W. S. Lee, P. S. Yu, and X. Li, "Partially supervised classification of text documents," in *Proc. Int. Conf. Machine Learning*, 2002, pp. 387–394.

[18] H. Yu, J. Han, and K. C.-C. Chang, "Pebl: Positive example based learning for web page classification using svm," in *Proc. ACM Int. Conf. Knowledge Discovery and Data Mining*, 2002, pp. 239–248.

[19] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.

[20] C. Brodley and M. Freidl, "Identifying mislabeled training data," *J. Artif. Intell. Res.*, vol. 11, pp. 131–167, 1999.

[21] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *Proc. Eur. Conf. Computational Learning Theory*, 1995, pp. 23–37.

[22] N. D. Lawrence and B. Scholkopf, "Estimating a kernel fisher discriminant in the presence of label noise," in *Proc. Int. Conf. Machine Learning*, 2001, pp. 306–313.

[23] R. P. W. Duin, D. de Ridder, and D. M. J. Tax, "Experiments with a featureless approach to pattern recognition," *Pattern Recognit. Lett.*, vol. 18, no. 11–13, pp. 1159–1166, Nov. 1997.

[24] E. Pezkalska and R. P. Duin, "Dissimilarity representations allow for building good classifiers," *Pattern Recognit. Lett.*, vol. 23, no. 8, pp. 943–956, Jun. 2002.

[25] G. P. Nguyen, M. Worring, and A. W. M. Smeulders, "Interactive search by direct manipulation of dissimilarity space," *IEEE Trans. Multimedia*, vol. 9, no. 7, pp. 1404–1415, Nov. 2007.

[26] E. Bruno, N. Moenne-Loccoz, and S. Marchand-Maillet, "Design of multimodal dissimilarity spaces for retrieval of video documents," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 9, pp. 1520–1533, Sep. 2008.

[27] V. J. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artif. Intell. Rev.*, vol. 22, pp. 85–126, 2004.

[28] D. Tao, X. Tang, X. Li, and X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 7, pp. 1088–1099, Jul. 2006.

[29] J. Platt, "Probabilistic outputs for support vector machines and comparison to regularized likelihood methods," in *Proc. Advances in Large Margin Classifiers*, 2000, pp. 61–74.

[30] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. New York: Wiley, 2004.

[31] H. Muller, W. Muller, D. M. Squire, S. Marchand-Maillet, and T. Pun, "Performance evaluation in content-based image retrieval: Overview and proposals," *Pattern Recognit. Lett.*, vol. 22, no. 5, pp. 593–601, Apr. 2001.

[32] SVM-Light [Online]. Available: http://svmlight.joachims.org/

[33] M. Grubinger, "Analysis and Evaluation of Visual Information Systems Performance," Ph.D. dissertation, Victoria Univ., Melbourne, Australia, 2007.