

University of Wollongong

Research Online

Faculty of Engineering and Information
Sciences - Papers: Part A

Faculty of Engineering and Information
Sciences

1-1-2002

A sub-vector weighting scheme for image retrieval with relevance feedback

Lei Wang

Nanyang Technological University, Singapore, leiw@uow.edu.au

Kap Luk Chan

Nanyang Technological University, Singapore

Xuejian Xiong

Nanyang Technological University, Singapore

Follow this and additional works at: <https://ro.uow.edu.au/eispapers>



Part of the [Engineering Commons](#), and the [Science and Technology Studies Commons](#)

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

A sub-vector weighting scheme for image retrieval with relevance feedback

Abstract

In this paper, a sub-vector weighting scheme is proposed for the case of small sample in image retrieval with relevance feedback. By partitioning a multi-dimensional visual feature vector to multiple sub-vectors, the singularity problem caused by small sample can be avoided by the lower dimensionality of the sub-vectors. Then the optimal weighting can be performed on these sub-vectors respectively and the similarity scores obtained are combined as the final score to rank the database images. Experimental results demonstrated that the proposed weighting scheme can significantly improve the efficacy of image retrieval with relevance feedback.

Keywords

vector, scheme, image, sub, retrieval, feedback, weighting, relevance

Disciplines

Engineering | Science and Technology Studies

Publication Details

Wang, L., Chan, K. Luk. & Xiong, X. (2002). A sub-vector weighting scheme for image retrieval with relevance feedback. *International Journal of Image and Graphics*, 2 (2), 199-213.

A Sub-vector Weighting Scheme for Image Retrieval with Relevance Feedback

Lei Wang, Kap Luk Chan, Xuejian Xiong

*School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore 639798
E-mail: P148373436@ntu.edu.sg*

Abstract

In this paper, a sub-vector weighting scheme is proposed for the case of small sample in image retrieval with relevance feedback. By partitioning a multi-dimensional visual feature vector to multiple sub-vectors, the singularity problem caused by small sample can be avoided by the lower dimensionality of the sub-vectors. Then the optimal weighting can be performed on these sub-vectors respectively and the similarity scores obtained are combined as the final score to rank the database images. Experimental results demonstrated that the proposed weighting scheme can significantly improve the efficacy of image retrieval with relevance feedback.

1. Introduction

Recently, relevance feedback [12, 11, 15, 7, 6, 2] is introduced into image retrieval to overcome the two obstacles below. One is the gap between the high-level concepts used by human and the low-level visual features used by computer, and the other is the subjectivity of human perception on visual content [12]. By learning from the images labelled as “positive” or “negative” in user feedback (i.e. taking these images as samples), it is possible to encapsulate the high-level concepts and the user’s perception embodied. Among the learning methods, feature component weighting is a commonly used one [4, 11, 1, 8]. By using the weights learned from the labelled images, the feature components in a similarity metric are weighted to simulate, to some extent, the high-level concepts and the user’s perception. The optimal weighting scheme has been derived in [4] and the corresponding similarity metric is essentially a Mahalanobis distance [3] in essence, in which a non-singular covariance matrix is needed. However, the number of the labelled images in user feedback is often small because a user will not recognize and label a great number of images in a retrieval process. In fact, this is what a retrieval system should do for the user. On the other hand, unfortunately, the

visual feature vectors used are multi-dimensional and some of them can be of the order of 10^2 [10]. It can be shown that at least $n + 1$ samples are needed to obtain a non-singular $n \times n$ covariance matrix. Obviously, the number of the labelled images obtained from relevance feedback is far less than what is required here.

Commonly, the method used to deal with the singularity problem is to ignore the mutual correlation among the feature components by assuming that these components are independent to each other [6, 11, 15, 9]. Thus, the original covariance matrix is replaced by a diagonal one and the singularity is avoided. Frankly, though the assumption of independence often deviates from the reality, it may be an option that has to be taken when the number of the labelled images is very small. This is because any estimation for the mutual correlation now will be a bad disturbance rather than a useful description. However, the independent weighting method should be reviewed when the situation changes.

Not as popular as the method above, another method using *pseudo-inverse matrix* is proposed in [4]. Essentially, it is identical to Principal Component Analysis (PCA) [3]. This method tries to avoid the singularity problem by reducing the original high-dimensional feature space to a lower one. A similar method, well known as the “Eigenface”, has been used successfully in facial image recognition [14]. However, the good performance of the “Eigenface” is founded on the similarity of facial images overall and the corresponding data can be represented well by a much lower subspace. Generally, the requirement cannot be satisfied in image retrieval because images are recognized by high-level semantics and the appearances of two images may be different entirely even if they come from the same class. Moreover, the lower-dimensional subspace cannot be estimated reliably because of the case of small sample.

To overcome the drawbacks of the methods above, this paper proposes a sub-vector weighting scheme. In this scheme, a multi-dimensional visual feature vector is partitioned to multiple sub-vectors. Because of the low dimensionality, the singularity of the covariance matrix for each sub-vector can be avoided when there is sufficient but small

number of the labelled images. Then the optimal weighted similarity metric can be performed on each sub-vector. The multiple similarity scores obtained are combined as the final score to rank the database images to get the retrieval result. Experimental results demonstrated that the proposed method can significantly improve the efficacy of image retrieval. The rest of the paper is organized as follows. In section 2, the related work is presented. In section 3, the sub-vector weighting scheme is proposed. Section 4 presents the experimental results and discussions. Finally, concluding remarks are given in section 5.

2. Related work

Assume that m image samples are fed back by a user and the corresponding n -dimensional visual feature vectors are $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m$, where $\mathbf{X}_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T$. By arguing that the weighting scheme based on the *pseudo-inverse matrix* below can only get around the singularity problem in theoretical sense, Rui *et. al.* in [9] proposed to use the standard deviation (STD) based weighting scheme instead when the number of image samples is smaller than the dimension of the feature vector, \mathbf{X} . In this scheme, the weighting matrix is assumed to be a diagonal one and the mutual correlation among the feature components is discarded completely. The similarity metric of this scheme can be shown as

$$d_i = (\mathbf{X}_i - \bar{\mathbf{X}})^T \mathbf{C}_d^{-1} (\mathbf{X}_i - \bar{\mathbf{X}}) \quad (1)$$

where $\bar{\mathbf{X}} = \frac{1}{m} \sum_{i=1}^m \mathbf{X}_i$ and $\mathbf{C}_d = \text{diag}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2\}$ where σ_i ($1 \leq i \leq n$) is the standard deviation of the i_{th} feature components of \mathbf{X} . It can be seen that the singularity is avoided by using the diagonal covariance matrix \mathbf{C}_d to replace the original matrix \mathbf{C} .

Ishikawa *et. al.* in [4] derive the optimal weighting scheme. It defines a concept of “goodness”, v , that is proportional to the degree of an image’s similarity to query. Considering that the images with high goodness scores judged by users should have small distances to the query, it minimizes a sum of the goodness-weighted similarity distances of the m images to the query, and obtains the optimal weighted similarity metric as

$$d_i = K(\mathbf{X}_i - \bar{\mathbf{q}})^T \mathbf{C}^{-1} (\mathbf{X}_i - \bar{\mathbf{q}}) \quad (2)$$

where

$$\begin{aligned} q_k &= \frac{\sum_{i=1}^m v_i x_{ik}}{\sum_{i=1}^m v_i} \\ c_{jk} &= \sum_{i=1}^m v_i (x_{ik} - q_k)(x_{ij} - q_j) \end{aligned} \quad (3)$$

Because $K = \det(\mathbf{C})^{\frac{1}{n}}$ is independent of images, removing it will not change the rank of these images. Moreover,

It can be found that \mathbf{C} is just a covariance matrix of different magnitude if the common bi-level (1-“positive” and 0-“negative”) goodness is used. Hence, the optimal weighted similarity metric is just a Mahalanobis distance in essence. When the number of the image samples is smaller than $n + 1$, the singularity of \mathbf{C} will be encountered.

The *pseudo-inverse matrix* is proposed in [4] to deal with the singularity of the matrix \mathbf{C} . Considering the covariance matrix \mathbf{C} is symmetric, a decomposition of \mathbf{C} is done as

$$\mathbf{C} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^T \quad (4)$$

where $\mathbf{\Gamma}$ is an orthogonal matrix and $\mathbf{\Lambda}$ is a diagonal one. The columns of $\mathbf{\Gamma}$ are the eigenvectors of \mathbf{C} and the diagonal elements of $\mathbf{\Lambda}$ are the corresponding eigenvalues. When only m ($m < n + 1$, recall that n is the dimension of \mathbf{X}) image samples are available, $\mathbf{\Lambda}$ has the form as

$$\mathbf{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_{m-1}, 0, \dots, 0\} \quad (5)$$

Define

$$\mathbf{C}^+ = \mathbf{\Gamma} \mathbf{\Lambda}^+ \mathbf{\Gamma}^T \quad (6)$$

where $\mathbf{\Lambda}^+ = \text{diag}\{\lambda_1^{-1}, \dots, \lambda_{m-1}^{-1}, 0, \dots, 0\}$ and it is called the *pseudo-inverse matrix* of \mathbf{C} . It can be found that the original n -dimensional feature space is reduced to a subspace with the dimension of $(m - 1)$ and the singularity is avoided.

3. The proposed sub-vector weighting scheme

3.1. The basic idea

Without loss of generality, the common bi-level goodness is used and the matrix \mathbf{C} turns out to be the covariance matrix $\mathbf{\Sigma}$. From m ($m < n + 1$) positive examples, the unbiased estimation of the covariance matrix of \mathbf{X} can be shown as

$$\hat{\mathbf{\Sigma}} = \frac{1}{m-1} \sum_{i=1}^m (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T \quad (7)$$

and it can be rewritten as

$$\hat{\mathbf{\Sigma}} = \frac{1}{m-1} \mathbf{A} \mathbf{A}^T \quad (8)$$

where $\mathbf{A}_{n \times m} = [(\mathbf{X}_1 - \bar{\mathbf{X}}) \ (\mathbf{X}_2 - \bar{\mathbf{X}}) \ \dots \ (\mathbf{X}_m - \bar{\mathbf{X}})]$. It is easy to prove that $\text{rank}(\mathbf{A}) = m - 1$ when $\mathbf{X}_1, \dots, \mathbf{X}_m$ are independent to each other. Then

$$\begin{aligned} \text{rank}(\hat{\mathbf{\Sigma}}) &= \text{rank}(\mathbf{A} \mathbf{A}^T) \\ &\leq \min [\text{rank}(\mathbf{A}), \text{rank}(\mathbf{A}^T)] \\ &= \text{rank}(\mathbf{A}) = m - 1 \end{aligned} \quad (9)$$

Equation (9) indicates that the rank of $\hat{\Sigma}$ is no more than $m - 1$. Hence, to obtain a non-singular $n \times n$ covariance matrix, $n + 1$ samples are necessary at least.

Fortunately, each k -order principal sub-matrix of the $\hat{\Sigma}$, $\hat{\Sigma}_k$, can be viewed as the covariance matrix for a sub-vector including the corresponding k feature components (Note that a principal sub-matrix of a matrix \mathbf{B} is the square sub-matrix whose diagonal lies on the diagonal of \mathbf{B}). Generally, the covariance matrix for a k -dimensional sub-vector will not be singular when $k \leq m - 1$. Simultaneously, to get, as possible as it can, a reliable estimation of $\hat{\Sigma}_k$, a low enough dimension of $k = 2$ is used for this sub-vector here. Thus, the mutual correlation among the feature components in the same sub-vector can be kept well and the data need not to be projected to an unreliable subspace with a much lower dimension. However, the correlation between each pair of the sub-vectors has to be discarded still because of the fact of small sample.

3.2. Constructing the sub-vectors

As mentioned above, the dimensions of all sub-vectors are set as $k = 2$ (Please note that k is still used in the following paper instead of 2 for the convenience of representation). Generally, let $n = tk + k'$, where t and k' ($k' = n \bmod k$) are the quotient and the remainder, respectively. Then the n -dimensional feature vector \mathbf{X} will be partitioned to $t + 1$ sub-vectors named $\mathbf{S}_1, \dots, \mathbf{S}_{t+1}$. Referring to generalized correlation analysis, the correlation coefficient between the two sub-vectors of \mathbf{S}_i and \mathbf{S}_j , R_{ij} , can be shown as

$$R_{ij} = 1 - \frac{|\Sigma|}{|\Sigma_{\mathbf{S}_i \mathbf{S}_i}| |\Sigma_{\mathbf{S}_j \mathbf{S}_j}|} \quad (10)$$

where $\Sigma_{\mathbf{S}_i \mathbf{S}_i}$ and $\Sigma_{\mathbf{S}_j \mathbf{S}_j}$ are, respectively, the covariance matrices of \mathbf{S}_i and \mathbf{S}_j while the Σ is the covariance matrix for the vector of $[\mathbf{S}_i^T, \mathbf{S}_j^T]^T$.

Then the optimal partitioning, θ^* , can be shown as

$$\theta^* = \arg \min_{\theta \in \Theta} \left(\sum_{i=1}^{t+1} \sum_{j>i}^{t+1} R_{ij} \right) \quad (11)$$

where Θ is the set including all possible partitions. Easily, it can be found that the number of the possible partitions will be huge even if n is not a large number. In this case, searching exhaustively for the optimal partitioning is unrealistic in real-time retrieval. According to matrix theory, there is

$$R_{ij} = 0 \iff |\Sigma'| = |\Sigma_{ii}| |\Sigma_{jj}| \iff \Sigma_{ij} = \mathbf{O} \quad (12)$$

This equation indicates that to minimize R_{ij} , the mutual correlation among the feature components in different sub-vectors should be as weak as possible to make $\Sigma_{ij} \rightarrow \mathbf{O}$.

That's to say, the pair of strongly correlated feature components should be combined into the same sub-vector as possible as they can. Based on this, a sub-optimal partitioning scheme is proposed as follows.

1. By using the m positive examples, calculate the correlation coefficient between the feature components of x_i and x_j , where $1 \leq i < j \leq n$.
2. Combine the pair of feature components having the maximal correlation coefficient as the first sub-vector.
3. From the remainder feature components, combine the pair having the maximal correlation coefficient as the second sub-vector.
4. Redo step. 3 until t pairs of the components are constructed.
5. Label the obtained component pairs as the sub-vectors of $\mathbf{S}_1, \dots, \mathbf{S}_t$ when n is even or as $\mathbf{S}_1, \dots, \mathbf{S}_{t+1}$ when n is odd.

It can be found that the proposed sub-optimal scheme only concerns the maximal correlation and does not balance among these correlations of R_{ij} to minimize the sum of them. However, the sub-optimal scheme is a compromise to reduce the computational load so that it can be used for real-time retrieval in practice. Note that the response time is as important as the efficacy for a retrieval algorithm.

3.3. The weighted similarity metric

Once the sub-vectors are decided, \mathbf{X} can be shown as

$$\mathbf{X} = [x_1, x_2, \dots, x_n]^T = [\mathbf{S}_1^T, \mathbf{S}_2^T, \dots, \mathbf{S}_{t+1}^T]^T \quad (13)$$

Thus, the covariance matrix Σ can be shown as

$$\Sigma = \text{diag}\{\Sigma_1, \Sigma_2, \dots, \Sigma_{t+1}\} \quad (14)$$

where Σ_i is the covariance matrix of \mathbf{S}_i and it can be estimated from the m positive examples. The premise of $k \leq m - 1$ ensures that Σ_i will not be singular.

According to matrix theory, given $\mathbf{A}_{m \times m}$ and $\mathbf{B}_{n \times n}$ are two non-singular matrices, there is

$$\text{diag}\{\mathbf{A}, \mathbf{B}\}^{-1} = \text{diag}\{\mathbf{A}^{-1}, \mathbf{B}^{-1}\} \quad (15)$$

Then the inverse matrix of Σ can be shown as $\Sigma^{-1} = \text{diag}\{\Sigma_1^{-1}, \Sigma_2^{-1}, \dots, \Sigma_{t+1}^{-1}\}$. Referring to the optimal weighted similarity metric in [4], the similarity metric of the proposed weighting scheme can be derived as

$$\begin{aligned} d_i &= (\mathbf{X}_i - \bar{\mathbf{X}})^T \Sigma^{-1} (\mathbf{X}_i - \bar{\mathbf{X}}) \\ &= \sum_{j=1}^{t+1} (\mathbf{S}_{i,j} - \bar{\mathbf{S}}_j)^T \Sigma_j^{-1} (\mathbf{S}_{i,j} - \bar{\mathbf{S}}_j) \end{aligned} \quad (16)$$

where $\mathbf{S}_{i,j}$ is the j_{th} sub-vector of \mathbf{X}_i and $\bar{\mathbf{S}}_j$ is the mean of \mathbf{S}_j estimated from the m positive examples.

It can be seen that the new similarity metric is the sum of the Mahalanobis distances corresponding to these sub-vector respectively. This similarity metric is more effective than the standard deviation weighted metric because it concerns the mutual correlation of feature components. Meanwhile, it is also more practical than the weighting based on the *pseudo-inverse matrix*.

4. Experimental Results

In this experiment, the proposed weighting scheme based on sub-vectors (WSV in short) will be compared with the standard deviation based weighting (WSTD in short) and the weighting essentially based on Principal Component Analysis (WPCA in short).

It is well-known that the effectiveness of the Mahalanobis distance, the basis of the new similarity metric above, is based on the assumption of Gaussianity of the data. Hence, the comparison will be carried out, respectively, on a real image database and an artificial database in which each of the classes conforms a Gaussian distribution strictly. The real image database includes 600 general color images composed from *VisTex* of MIT and *Corel* Stock Photos. Every 100 images form a class and six classes are founded based on high-level semantics. The classified database provides the ground truth to evaluate the various methods. A perceptually uniform color space, *CIE - Lab*, is used to represent general color images. Based on the color space, two visual feature vectors are defined. One is the color feature vector known as color moments [13]. It includes the mean, variance, and skewness of the values of an image in L , a and b axes. The dimension of color moments is 3×3 (9 in total). The other is the Gabor based texture feature [5]. By utilizing a bank of Gabor filters (6 orientations and 4 scales), the first and second order moments of the values of the Gabor filtered image in L , a , and b axes are extracted to form a feature vector. This feature vector has a high dimension of $6 \times 4 \times 2 \times 3$ (144 in total).

The artificial database is constructed based on the real image database and they share the same structure. The feature of color moment above is also used for the artificial database. By calculating the means and covariance matrices of this feature vector for the six classes of the real image database respectively, six Gaussian distributions are obtained correspondingly. From each Gaussian distribution, 100 samples are produced randomly to represent the 100 feature vectors of the color moments of the images in the same class. Thus, the color feature based artificial database is constructed. Similarly, the texture feature based artificial database is constructed. However, the dimension of the used texture feature vector here is not 144 above but 48 to avoid

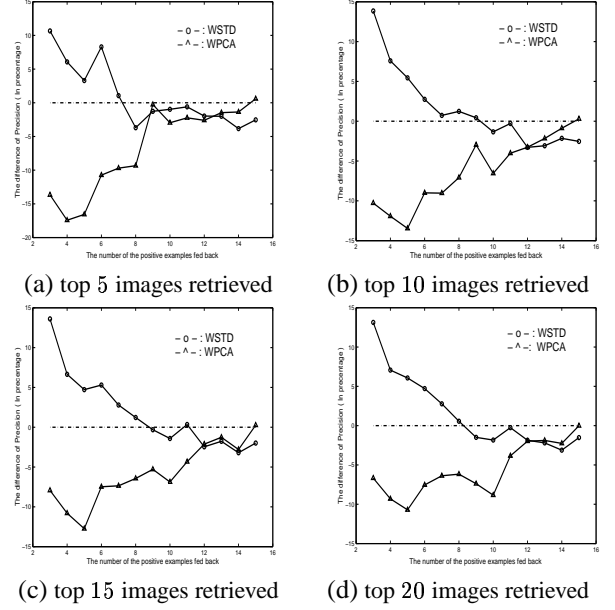


Figure 1. The comparison results of color feature on the artificial database

singularity because there is only 100 images in each class in the real image database. The 48-dimensional feature vector represents the first and second order moments of the values of the Gabor filtered image in L axis.

The *Precision* is used to measure the efficacy of a retrieval. It is defined as

$$Precision = \frac{Number_Retrieved_Positive}{Number_Total_Retrieved} \quad (17)$$

Let P_{WSTD} , P_{WPCA} , and P_{WSV} be the *Precision* of the methods of WSTD, WPCA, and WSV, respectively. To compare the proposed weighting scheme with other methods, the differences of the three *Precision* values are defined as follows.

$$\begin{aligned} \Delta P_{WSTD} &= P_{WSTD} - P_{WSV} \\ \Delta P_{WPCA} &= P_{WPCA} - P_{WSV} \end{aligned} \quad (18)$$

If ΔP_{WSTD} or ΔP_{WPCA} is negative, it means that an improvement is gained by the proposed weighting scheme.

To simulate the case of small sample, especially, the case in the first time feedback, the number of positive examples, m , is limited from 3 to 20.

The process of the experiment can be described as follows.

1. Use an image in class i as the example of query and launch a retrieval.

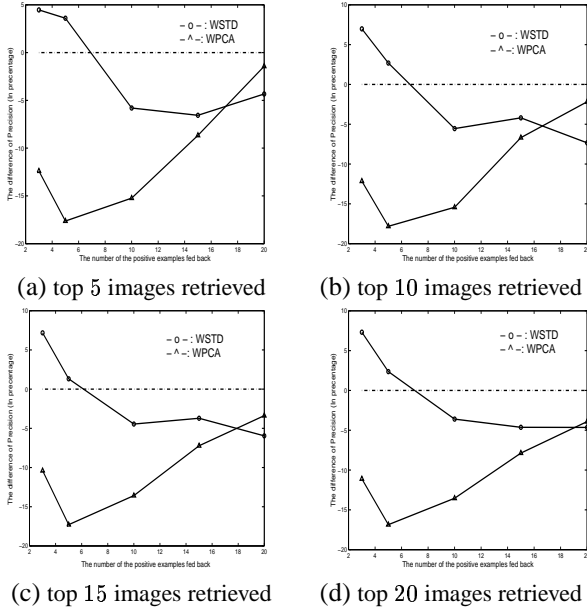


Figure 2. The comparison results of texture feature on the artificial database

2. By using the Euclidean distance as the initial similarity metric, the database images are ranked in a descending order of the corresponding distances.
3. Referring to the rank above, the first m images that share the same class number with the query are used to simulate the positive examples fed back by the user in the process of relevance feedback.
4. By using the m positive examples, the methods of WSV, WSTD, and WPCA are used respectively to retrieve the other positive images from the database.
5. To accumulate the statistics, redo step. 1 to 4 until all images in class i have been used as query. For each method, the average of the *Precision* values for all retrievals is calculated as the *Precision* for this method to class i when m positive examples are used.
6. Redo step. 1 to 5 for class i ($1 \leq i \leq 6$). For each method, the average of the *Precision* values of the six classes is calculated as *Precision* for this method to the whole database when m positive examples are used.
7. Change m from 3 to 20 and redo step. 1 to 6 to obtain the *Precision* of each method under different numbers of positive examples.
8. Compare the obtained *Precision* curves of the three methods.

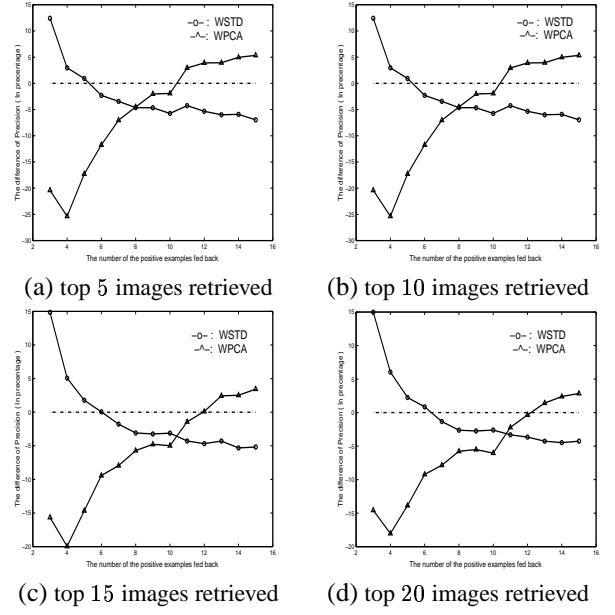


Figure 3. The comparison result of color feature on real image database

4.1. The experimental results and discussions

Figure 1 and 2 show the variations of ΔP_{WSTD} and ΔP_{WPCA} with the number of the positive examples available. The line labelled as *WSTD* shows ΔP_{WSTD} and the line of *WPCA* shows the ΔP_{WPCA} . Figure 1 is the result of using the feature of color moments. The four sub-figures of Figure 1(a), (b), (c), and (d) shows the cases for the top 5, 10, 15, and 20 images respectively. By comparing the two solid curves with the dashed line indicating zero value, it can be seen that the performance of WSV is consistently the best one when the number of positive examples is above 8. When the number is smaller than 8, the performance of the proposed method is worse than WSTD because the estimation for each sub-vector's covariance matrix is so unreliable that the obtained estimation becomes a negative disturbance rather than a useful information for the proposed weighting scheme. When the number is larger than 15, the method based on PCA shows the best performance because the singularity problem is not there any more and the estimation for the covariance matrix can be more and more reliable with more positive examples. Because the dimensions of the color feature is only 9, the duration of showing the best performance is short for the proposed method and the improvement over the method of WSTD is no more than 5%. However, as shown in Figure 2, when the 48-dimensional texture feature vector is used, the proposed weighting scheme shows the best performance consistently

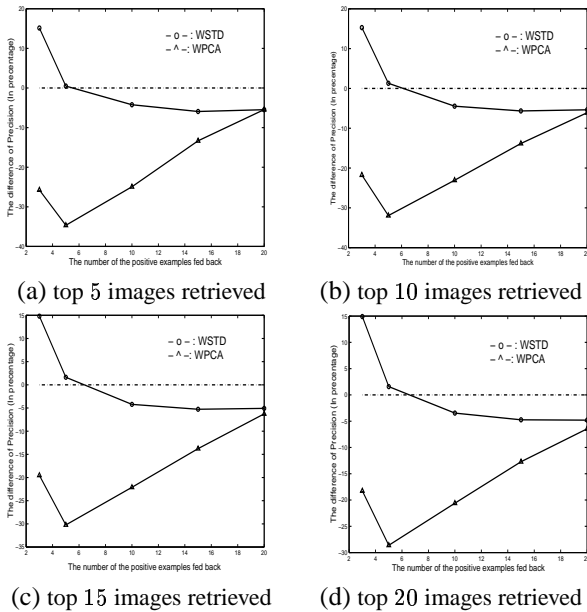


Figure 4. The comparison result of texture feature on real image database

as long as the number of positive examples is larger than 7, and the improvement over WSTD can reach 5% – 10%. With the increasing on the number of the positive examples, a rapid improvement can be seen on the performance of WPCA and it exceeds the method of WSTD when the number is still much smaller than the dimension of 48. Again, it is shown that the method of WSTD will not be effective any more when more image samples become available.

It can be seen that the proposed weighting scheme achieves good performance on the artificial database where the Gaussianity of the data is satisfied. A similar finding can be obtained from the results for the real image database shown in Figure 3 and Figure 4.

5. Conclusion

In this paper, a sub-vector weighting scheme is proposed for image retrieval with relevance feedback. By partitioning a multi-dimensional feature vector to multiple sub-vectors, this scheme tries to represent the original data as effectively as possible by using the limited image samples in user feedback. A new weighted similarity metric is obtained correspondingly. The experimental results show the better performance achieved by the proposed weighting scheme when it is applied in the case of small sample. In the future work, further verification will be carried out on a larger image database.

References

- [1] S. Aksoy, R. M. Haralick, F. A. Cheikh, and M. Gabbouj. A weighted distance approach to relevance feedback. *Proceedings of 15th International Conference on Pattern Recognition*, pages 812–815, 2000.
- [2] I. J. Cox, M. L. Miller, S. M. Omohundro, and P. N. Yianilos. Pichunter: Bayesian relevance feedback for image retrieval. *Proceedings of the 13th International Conference on Pattern Recognition, 1996*, 3:361–369, 1996.
- [3] R. O. Duda, D. G. Stork, and P. E. Hart. Pattern classification(second edition). *John Wiley and Sons*, 2001.
- [4] Y. Ishikawa, R. Subramanya, and C. Faloutsos. Mindreader: Querying databased through multiple examples. *Proc. of the 24th VLDB Conference (New York)*, pages 218–227, 1998.
- [5] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8), August 1996.
- [6] C. Meilhac and C. Naster. Relevance feedback and category search in image retrieval. *Proceedings of IEEE International Conference on Multimedia Computing and Systems*, pages 512–517 vol.1, 1999.
- [7] T. P. Minka and R. W. Picard. Interactive learning using a “society of models. *Technical Report 349, M.I.T. Media Laboratory Perceptual Computing Section*, 1995.
- [8] K. Porkaew and S. Mehrotra. Query reformulation for content based multimedia retrieval in mars. *Proceedings of IEEE International Conference on Multimedia Computing and Systems*, pages 747–751 vol.2, 1999.
- [9] Y. Rui and T. S. Huang. Optimizing learning in image retrieval. *Proc. of IEEE Int Conf on Computer Vision and Pattern Recognition (CVPR), Hilton Head, SC*, 2000.
- [10] Y. Rui, T. S. Huang, and S.-F. Chang. Image retrieval: Past, present, and future. *Int Symposium on Multimedia Information Processing , Dec 11-13, 1997, Taipei, Taiwan*, 1997.
- [11] Y. Rui, T. S. Huang, and S. Mehrotra. Content-based image retrieval with relevance feedback in mars. *Proc. of IEEE Intrenational Conference on Image Processing, Santa Barbara, California, USA*, 1997.
- [12] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra. Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Trans. on Circuits and Systems for Video Technology*, 8(5):644–655, 1998.
- [13] M. Stricker and M. Orengo. Similarity of color images. *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1995.
- [14] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. *Proceedings of IEEE Computer Society Conference Computer Vision and Pattern Recognition*, pages 586–591, 1991.
- [15] N. Vasconcelos and A. Lippman. Learning from user feedback in image retrieval systems. *Proc. of NIPS’99, Denver, Colorado*, 1999.