Relevance feedback-based building recognition

Jing Li and Nigel M. Allinson
Imaging Engineering Laboratory, Department of Electronic and Electrical Engineering, University of Sheffield, Mappin Street, Sheffield, S1 3JD, U.K.

ABSTRACT

Building recognition is a nontrivial task in computer vision research which can be utilized in robot localization, mobile navigation, etc. However, existing building recognition systems usually encounter the following two problems: 1) extracted low level features cannot reveal the true semantic concepts; and 2) they usually involve high dimensional data which require heavy computational costs and memory. Relevance feedback (RF), widely applied in multimedia information retrieval, is able to bridge the gap between the low level visual features and high level concepts; while dimensionality reduction methods can mitigate the high-dimensional problem. In this paper, we propose a building recognition scheme which integrates the RF and subspace learning algorithms. Experimental results undertaken on our own building database show that the newly proposed scheme appreciably enhances the recognition accuracy.

Keywords: Building recognition, gist features, dimensionality reduction, relevance feedback, SVM

1. INTRODUCTION

Building recognition aims to distinguish different buildings in a large-scale image database. It is an intra-class classification task that can be utilized in many practical application areas, e.g., architectural design, building labelling in videos, robot vision for localization [36], and mobile device navigation [1][12]. However, this task is more challenging compared to general object recognition since building images may be taken from different viewpoints, under different lighting conditions, or suffer from occlusion from trees, moving vehicles, other buildings or themselves. As a result, little attention has been paid to this specific recognition task. In [13], perceptual grouping in a hierarchical way was applied to explore the semantic relationships among low-level visual features for content-based image retrieval (CBIR) for buildings. Consistent line clusters, a type of mid-level feature, were put forward in [20] for building recognition in CBIR. After extracting colour, orientation, and spatial information for each line segment, they were grouped into different consistent clusters, and the intra-cluster and inter-cluster relationships can be utilized to recognize different buildings. Zhang and Kosecká [40] proposed a building recognition system based on vanishing point detection and localized colour histograms. Because of the fast indexing step using localized colour histograms, this method achieved some improvement in efficiency. Hutchings and Mayol [12] applied the Harris corner detector [8] to extract interest points for matching buildings in the world space for mobile device.

Nevertheless, the building recognition approaches mentioned above suffer from some limitations, namely i) they are based on the detection of low level visual features, e.g., line segments, vanishing points, etc., which cannot reveal the proper semantic concepts of images and thus limits their representational performance; and ii) recognition is conducted on pairs of raw high-dimensional feature vectors, resulting in high computational cost and memory requirements.

To this end, we propose a building recognition scheme to address the above problems by integrating the relevance feedback (RF) and dimensionality reduction methods.

Since there is a semantic gap [25] between the low level visual features and high level image concepts, the retrieval performance based only on low level features is not satisfactory and it is very important to fully take into account the interaction between users and search engines. RF serves as a means to bridge this gap and has shown its superiority in terms of both effectiveness and efficiency. Traditional RF methods in CBIR include the following two steps [25][17]: i) when retrieved images are returned to a user, some relevant and irrelevant images are labelled by the user as positive and negative samples, respectively; and ii) the retrieval system refines the retrieved results based on these labelled samples. These two steps are conducted iteratively until the user is satisfied with the presented results.

Visual Communications and Image Processing 2010, edited by Pascal Frossard,
Houqiang Li, Feng Wu, Bernd Girod, Shipeng Li, Guo Wei, Proc. of SPIE Vol. 7744,
77440A · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.863191

Dimensionality reduction can make data more compact and is able to alleviate the high-dimension problem. Subspace learning-based dimensionality reduction, which finds a projection that reduces the high dimensional feature space to a lower dimensional subspace, is gaining more popularity in data mining and computer vision applications. A number of subspace learning methods, widely applied in biometrics [28][38][39] and multimedia information retrieval [9][29][30], have been proposed in the literature. Among these, conventional linear dimensionality reduction techniques [15][22] and manifold learning algorithms [2][5][10][24][32] dominate

The proposed scheme consists of the following three stages: i) feature representation; ii) dimensionality reduction; and iii) relevance feedback-based recognition. Features that are biologically present in the human visual perception are extracted using saliency and gist model [26], where the saliency model is constructed by extracting visual information from a raw image and the gist model is based on the extracted visual features. In detail, visual features are extracted at multi-scales and a set of feature maps is created for each image, resulting in the saliency model. After that, the gist model is constructed by dividing each feature map into a number of sub-regions and describing each map by a gist feature. In order to reduce the computational cost while preserving most of the discriminative information for recognition, dimensionality reduction is conducted to reduce the original higher dimensional feature vectors to a much lower dimensional feature space. Afterwards, building recognition is implemented by performing RF on our own building database [41], which incorporates various challenges including scaling, rotation, illumination changes, viewpoint changes, and camera-shake.

The organization of this paper is that we describe the biological-based features representation in Section 2; dimensionality reduction algorithms are introduced in Section 3; the adopted RF algorithm is briefly reviewed in Section 4; recognition performance on our own database is detailed in Section 4; while Section 5 concludes.

2. FEATURE REPRESENTATION

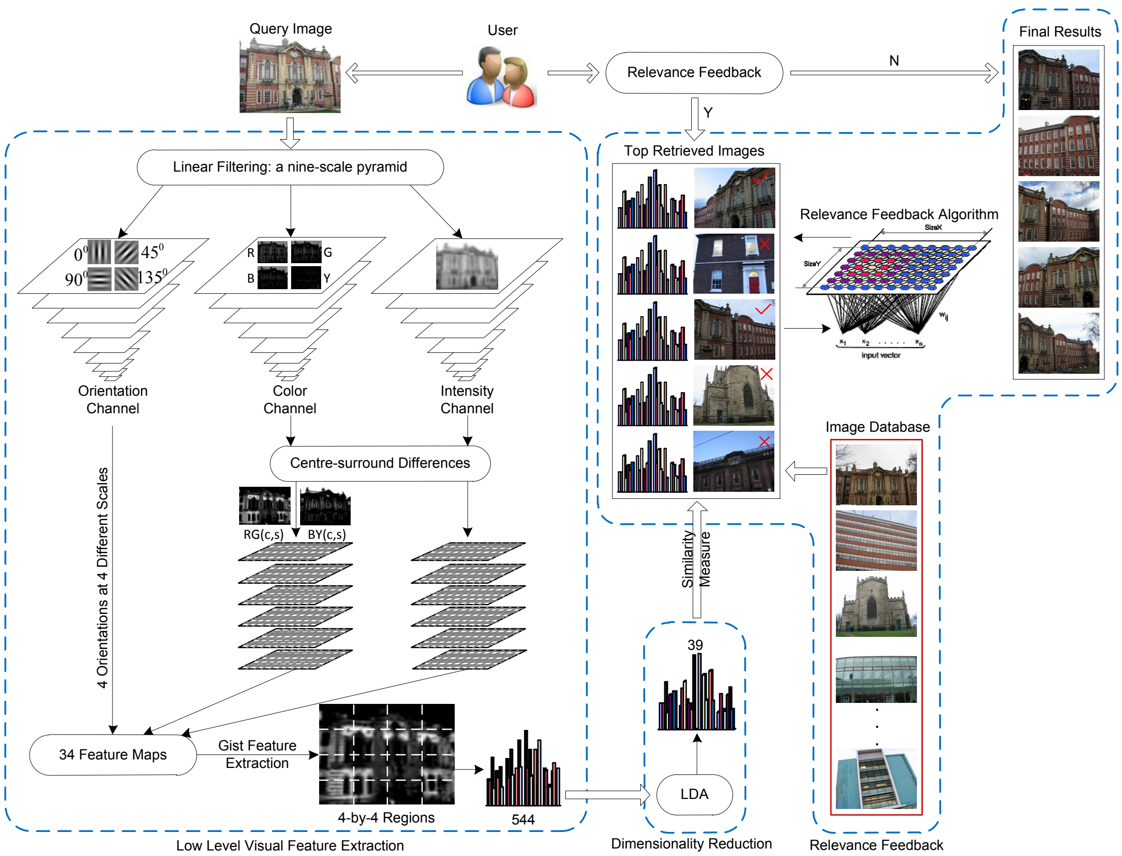

Figure 1. The proposed building recognition scheme.

Psychological research [34] supports the view that humans are able to grasp the holistic information in an image - also known as the gist [23] of a scene - (e.g., indoor or outdoor, approximate locations, dominative colours) by glancing at it for just a few of seconds. Inspired by this biological concept in [26], we utilize gist features as well as low-level visual features for building recognition that are known to provide more discriminative information for this type of recognition task. The key idea of feature representation is to first construct saliency feature maps, followed by generating a gist feature for each of them.

Saliency feature maps are constructed based on low-level visual features, including both global and local features, which are extracted in parallel. Global features are obtained by extracting intensity and colour information at different scales; local features are acquired by generating Gabor features [6][7] at several different scales and orientations. Although the features mentioned above have been previously utilized for other recognition tasks [27], it is the first time they are combined together for building recognition. After low-level feature extraction, a gist feature is extracted for each saliency map. To integrate low-level visual features with gist features, three major steps are conducted: i) linear filtering; ii) visual feature extraction; and iii) gist feature generation. Each step will be detailed in the following sections and the overall recognition scheme is represented in Figure 1.

2.1 Linear Filtering

Each input image $I(\vec{p})$ with $\vec{p}=[x,y]^T$, is linearly filtered to give a Gaussian pyramid of nine scales $I(\vec{p};\sigma)$ ($\sigma=0,...,8$). Afterwards, the centre-surround operation [14], widely used in modelling the receptive fields in the human visual system, is conducted by computing pixel differences across scales:

$$I_{c,s}(\vec{p})=\left|I(\vec{p};c)-I(\vec{p};s)\right|, \quad (1)$$

where $I(\vec{p};c)$ denotes a pixel at a centre scale $c=2,3,4$, $I(\vec{p};s)$ is its corresponding pixel at a surround scale $s=c+d$ with $d=3,4$, and $I(\vec{p};s)$ is interpolated to the centre scale so that pixel difference maps are obtained.
For simplicity, the centre-surround operation is reformulated as

$$I_{c,s}=\left|I_c \Theta I_s\right|, \quad (2)$$

It should be noted that linear filtering is a prerequisite step for all visual feature extractions, but the centre-surround is only performed for intensity and colour information extraction, which creates on centre-surround difference maps of $I_{2,5}$, $I_{2,6}$, $I_{3,6}$, $I_{3,7}$, $I_{4,7}$, and $I_{4,8}$

2.2 Visual feature extraction

Generally, there are many robust types of features for object recognition, such as colour, shape, texture, intensity, motion, etc. Applying more types of features may permit greater accuracy for classification; on the other hand, it means more computational complexity is required. To balance the accuracy and computational cost, only three types of visual features are adopted for the building recognition task, i.e., intensity, colour and orientation. This results in three main visual channels: intensity channel, colour channel, and orientation channel. To simplify the description, the intensity and colour information extraction are merged in a single sub-section.

2.2.1 Intensity and Colour

The combination of red (R), green (G), blue (B), and yellow (Y) can balance between reflectance and colour saturation. Three pairs of colour opponents [35], i.e., red-green (RG), blue-yellow (BY), and dark-bright intensity, are computed from them.

The intensity channel is computed by

$$I=(r+g+b)/3, \quad (3)$$

and R, G, B, Y colour channels are also mapped from the r, g, b outputs [14]:

$$
\begin{aligned}
R &= r-(g+b)/2 \\
G &= g-(r+b)/2 \\
B &= b-(r+g)/2 \\
Y &= r+g-2(|r-g|+b)
\end{aligned}
\quad . \tag{4}
$$

Finally, three pairs of colour opponents are generated by performing the centre-surround difference [14]

$$
\begin{aligned}
RG_{c,s} &= \left|(R_c - G_c)\Theta(R_s - G_s)\right| \\
BY_{c,s} &= \left|(B_c - Y_c)\Theta(B_s - Y_s)\right| \\
I_{c,s} &= \left|I_c \Theta I_s\right|
\end{aligned}
\quad . \tag{5}
$$

As a result, 18 feature maps are obtained for intensity and colour.

2.2.2 Orientation

Gabor filters with four different scales and four orientations are utilized to extract the orientation information, making a total of 16 Gabor functions.

Research findings from cognitive psychology and psychophysics [21] suggest that Gabor filters [6][7] based on image decomposition are biologically relevant to human image understanding and recognition. Consequently, Gabor filters are appropriate for orientation information extraction within a purely computer vision context.

A Gabor filter [16] is the product of an elliptical Gaussian envelope and a complex plane wave, defined as:

$$
\psi_{s,d}(x,y) = \psi_{\vec{k}}(\vec{z}) = \frac{\|\vec{k}\|}{\delta^2} \cdot e^{-\frac{\|\vec{k}\|^2 \cdot \|\vec{z}\|^2}{2\delta^2}} \cdot \left[e^{i\vec{k}\cdot\vec{z}} - e^{-\frac{\delta^2}{2}} \right], \tag{6}
$$

where $\vec{z} = [x, y]$ is the variable in a spatial domain and $\vec{k}$ is the frequency vector, which determines the scale and orientation of Gabor filters, $\vec{k} = k_s e^{i\phi_d}$, where $k_s = k_{\max} / f^s$, $k_{\max} = \pi/2$, $f = 2$, $s = 0,1,2,3$, and $\phi_d = \pi d/8$, for $d = 0,2,4,6$.

After visual feature acquisition, there are 34 feature maps in total. Since both the centre-surround operation and Gabor filters are differential algorithms, the extracted visual features are robust to illumination changes.

2.3 Gist feature generation

A gist feature is generated from each feature map by dividing it into a 4×4 grid and then averaging the responses of pixels within each sub-region, resulting in a 16-dimensional feature vector for each feature map. Therefore, each image is represented by a 544-dimensional feature vector.

The gist feature for each image is expressed by

$$
G^k_{c,s}(i,j) = \frac{1}{16WH} \sum_{u=iW/4}^{(i+1)W/4-1} \sum_{v=jH/4}^{(j+1)H/4-1} \left[M^k_{c,s} \right](u,v), \tag{7}
$$

where $1 \le i, j \le 4$ for each map, W and H are the width and the height of an image, respectively, $1 \le k \le 34$, and $M^k_{c,s}$ denotes the k -th feature map at (c,s) . Since centre-surround operations are not conducted on Gabor features, $M_c \leftarrow M_{c,s}$ for gist feature generation on Gabor feature maps.

3. DIMENSIONALITY REDUCTION

To alleviate computational complexity, while preserving sufficient discriminative information in the subsequent recognition stage, the dimension of the original feature vectors is reduced from 544 to a much lower dimensional space, namely 39.

LDA [22] is a supervised learning algorithm that takes class label information into account. Given a set of labelled training examples, it finds the projection direction that maximizes the between-class scatter matrix while minimizing the within-class scatter matrix. It aims to separate examples from different classes far away while keeping those within the same class close to each other in the projected lower dimensional subspace.

Given that the original high dimensional data points $X = \{\vec{x}_1, \vec{x}_2, ..., \vec{x}_N\}$ in R^n are belonging to c classes, the formulation of LDA is:

$$U_{opt} = \arg\max_{U} \frac{U^T S_b U}{U^T S_w U}, \tag{8}$$

where $S_b = (1/N)\sum_{i=1}^{c} N_i (\vec{m}_i - \vec{m})(\vec{m}_i - \vec{m})^T$ is the between-class scatter matrix and $S_w = (1/N)\sum_{i=1}^{c}\sum_{j=1}^{N_i} (\vec{x}_{i;j} - \vec{m}_i)(\vec{x}_{i;j} - \vec{m}_i)^T$ is the within-class scatter matrix; in the i-th class, N_i is the number of data points, $\vec{x}_{i;j}$ represents the j^{th} example, and $m_i = (1/N_i)\sum_{j=1}^{n_i} \vec{x}_{i;j}$ is the mean value; $N = \sum_{i=1}^{c} N_i$ is the number of all training examples; and $m = (1/N)\sum_{i=1}^{c}\sum_{j=1}^{N_i} \vec{x}_{i;j}$ is the mean vector of the whole input data. The generalized eigenvalue problem is $S_b U = \lambda S_w U$, and the resulted lower dimensional subspace is spanned by $U = \{\vec{u}_1, \vec{u}_2, ..., \vec{u}_L\}$ ($L \le c-1$). Herein, the covariance matrix of all training examples is $S_t = (1/N)\sum_{i=1}^{c}\sum_{j=1}^{N_i} (\vec{x}_{i;j} - \vec{m})(\vec{x}_{i;j} - \vec{m})^T = S_b + S_w$.

4. RELEVANCE FEEDBACK

Support vector machines (SVMs) [37][4] are binary classifiers that maximize the margin between positive examples and negative examples. They have good generalization abilities and SVM-based relevance feedback algorithms [11][19][33] have been widely utilized in content-based image retrieval. However, they may lead to unsatisfactory performance when labelled feedback samples are limited. This is caused by: 1) SVM is unstable with a very small-sized training set since its optimal hyperplane is constructed by feedback samples; 2) SVM's optimal hyperplane may be unbalanced when the number of negative feedback samples is far more than that of positive feedback samples; and 3) overfitting may occur when the number of training examples is much less than the dimension of the feature vector.

Fortunately, asymmetric bagging and random subspace SVM (ABRS-SVM) [31] can alleviate the above mentioned problems and is adopted here as the relevance feedback scheme.

ABRS-SVM is the combination of asymmetric bagging-based SVM (AB-SVM) and random subspace SVM (RS-SVM) and thus inherits the merits of bagging [3] and random subspace method. In AB-SVM, bagging is implemented by bootstrapping on negative feedback samples and performing aggregation by the majority voting rule. It can solve the unstable and unbalance problems; while in RS-SVM, the random subspace method executes the bootstrapping in the feature space to deal with the overfitting problem.

In our proposed building recognition scheme, if relevance feedback is needed (Y), it will be conducted step by step; otherwise (N), final recognition results will be returned straightforward. The right part of Figure 1 shows the detail procedure of relevance feedback-based image retrieval system. In our case, the flow chat follows the 'Y' choice.

5. BUILDING RECOGNITION

In this section, we first introduce our self-constructed building database and then the experimental results conducted on this database are reported.

5.1 Database overview

Figure 2. (a)-(c) are sample images from the category 1, 10, and 38, respectively.

Experiments were carried out on our own image database [41], in which there are 40 buildings, i.e., 40 categories in total. All images in the database were taken around the University of Sheffield and Sheffield city centre using a Cannon G90 digital camera, including both stills and video clips. They were taken at different times on separate days. Different times cover early morning, noon, mid-afternoon, and early evening, resulting in highly variable lighting conditions and so making the building recognition task more challenging. Buildings include churches and a variety of modern buildings, such as exhibition halls. Images for each building were taken from different viewpoints, varying from three to nine views; by moving the camera from side to side, video clips are also obtained with multiple views. Furthermore, some videos were captured by walking from one side to another, which creates the additional challenge of movement/camera-shake. In summary, our building database incorporates a variety of challenges, i.e., rotation, scaling, variant lighting conditions, viewpoint changes, occlusions, and vibration. Video clips are of differing lengths, ranging from less than a 2 s to more than 10 s. We convert each clip into JPEG image format by sampling frames at rates from 3 to 10 frames/s, where the resolution of each frame is reduced to 160×120 pixels. Generally speaking, for each building, the number of images varies from 100 to 400, leading to unbalanced visual data within different building categories. To make sure all the images in the database are of the same size, each picture is resized to 160×120 pixels as well. In summary, there are 3,192 images. Sample images in the database are shown in Figure 2. As we can see, even humans cannot easily tell whether the images in the row (b) are from the same building or not. In addition, since the size of each image is small, as it is intended our developed system will have potential for mobile device navigation.

5.2 Experimental results

To explore whether RF enhances the recognition rates or not, the newly proposed scheme is evaluated on our image database, in which 100 images are randomly selected as the queries and RF is automatically performed. Each image is first represented by a 39-dimensional feature vector after low level visual feature extraction and dimensionality reduction. All images in the database are then sorted by a specified similarity measure. Within the top 40 returned images, the five most relevant images are labelled as positive feedback samples and five irrelevant images are marked as negative feedbacks. RF is conducted over nine iterations, for which precision and standard deviation (SD) are utilized to evaluate the performance. Precision is the percentage of the relevant images in the top N retrieved images, which describes the effectiveness of the RF algorithm; while SD serves as an error-bar to record its robustness. Both metrics are computed as the average values of the 100 query images. As we can see in Figure 3, after nine iterations, the retrieval average precision in top 10 results is up to 93%, which outperforms 85.25% without RF [18]. It demonstrates the effectiveness of RF in the building recognition task. In our experiments, the Gaussian kernel $K(\vec{x}, \vec{y}) = e^{-\rho\|\vec{x}-\vec{y}\|^2}$ with $\rho = 1$ (the default value in the OSU-SVM MatLab toolbox [42]) is chosen to evaluate the RF algorithm.

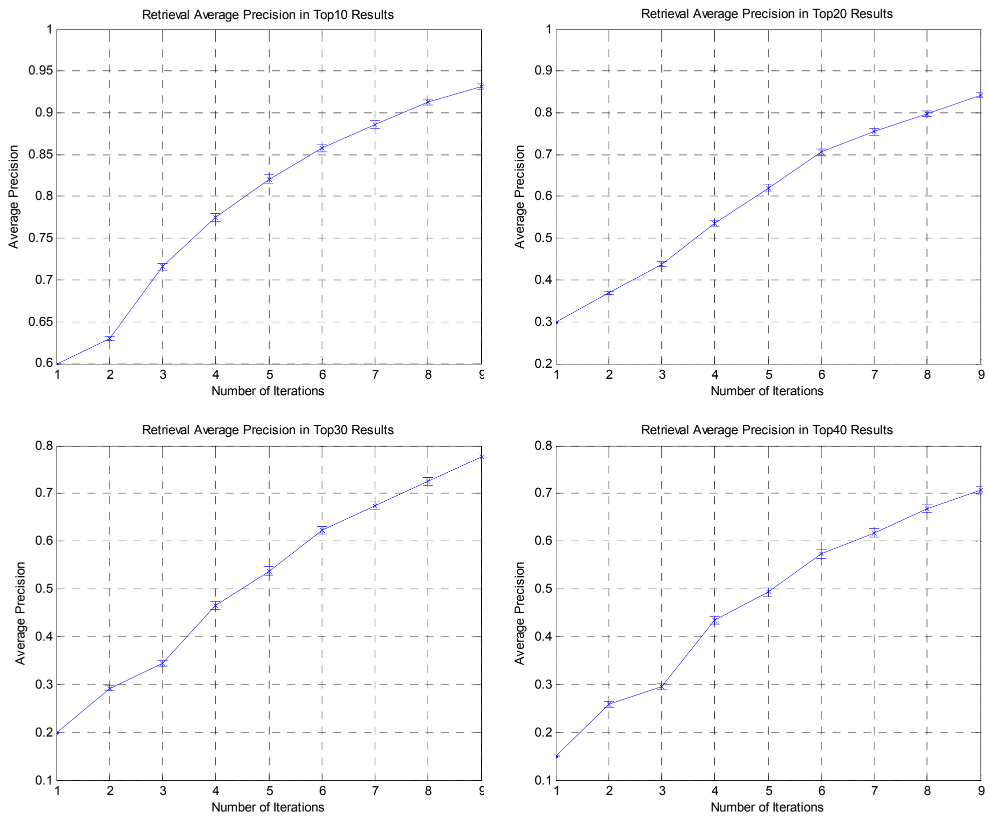

Figure 3. Recognition results in top 10, 20, 30, and 40 retrieved images, respectively. The ABRSVM is operated over 9 iterations.

6. CONCLUSIONS

This paper is the first to adopt the relevance feedback (RF) technique for the building recognition task, where a building recognition scheme is proposed by integrating RF and dimensionality reduction. Each image is represented by both saliency and gist models. Subspace learning-based dimensionality reduction is implemented to eliminate redundant information and thus alleviate the computational costs for the subsequent recognition process. Afterwards, a support vector machine (SVM)-based RF is implemented. Experiments are conducted on our own database. It is shown that RF enhances the recognition performance and the newly proposed scheme achieves satisfactory results.

REFERENCES

[1] Ali, H., Paar, G. and Paletta L., "Semantic Indexing for Visual Recognition of Buildings," Int'l Symposium on Mobile Mapping Technology, 28-31 (2007).

[2] Belkin, M. and Niyogi, P., "Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering," Advances in Neural Information Processing System, Vol. 14, 585-591 (2002).

[3] Breiman, L., "Bagging Predictors," Machine Learning, 24 (2), 123-140 (1996).

[4] Burges, J.C., "A Tutorial on Support Vector Machines for Pattern Recognition," Data Mining and Knowledge Discovery, 2(2), 121-167 (1998).

[5] Cai, D., He, X. and Han, J,, "Using Graph Model for Face Analysis," Technical Report, No. 2636, Department of Computer Science, University of Illinois at Urbana-Champaign, (2005).
[6] Daugman, J. G., "Two-Dimensional Spectral Analysis of Cortical Receptive Field Profile," Vision Research, Vol. 20, 847-856 (1980).
[7] Daugman, J. G., "Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two–Dimensional Visual Cortical Filters," J. Optical Society of America, 2(7), 1,160-1,169 (1985).
[8] Harris, C. and Stephens, M., "A Combined Corner and Edge Detector," Alvey Vision Conf., 147-151 (1988).
[9] He, X., Cai, D. and Han, J., "Learning a Maximum Margin Subspace for Image Retrieval," IEEE Trans. Knowledge and Data Engineering, 20(2), 189-201 (2008).
[10] He., X. and Niyogi, P., "Locality Preserving Projections," Proc. Conf. Advances in Neural Information Processing Systems, (2003).
[11] Hong, P., Tian, Q. and Huang, T.S., "Incorporate Support Vector Machines to Content-Based Image Retrieval with Relevant Feedback," Proc. IEEE Int'l Conf. Image Processing, 750-753 (2000).
[12] Hutchings, R. and Mayol-Cuevas, W., "Building Recognition for Mobile Devices: Incorporating Positional Information with Visual Features," CSTR-06-017, Computer Science, University of Bristol, (2005).
[13] Iqbal, Q. and Aggarwall, J.K., "Applying Perceptual Grouping to Content-based Image Retrieval: Building Images," Proc. IEEE Conf. Computer Vision and Pattern Recognition, Vol. 1, 42-48 (1999).
[14] Itti, L., Koch, C. and Niebur, E., "A Model of Saliency-based Visual Attention for Rapid Scene Analysis," IEEE Trans. Pattern Analysis and Machine Intelligence, 20(11), 1,254-1,259 (1998).
[15] Jolliffe, I. T., [Principal Component Analysis (2nd edition)], Springer, (2002).
[16] Lee, T. S., "Image Representation using 2D Gabor Wavelets," IEEE Trans. Pattern Analysis and Machine Intelligence, 18(10), 959-971 (2003).
[17] Li, J. and Allinson, N.M., "Long-term learning in content-based image retrieval," Int'l J. Imaging Systems and Technology, 18(2-3), 160-169 (2008).
[18] Li, J. and Allinson, N.M., "Subspace Learning-based Dimensionality Reduction in Building Recognition", Neurocomputing, 73(1-3), 324-330 (2009).
[19] Li, J., Allinson, N., Tao, D. and Li, X. "Multitraining Support Vector Machine for Image Retrieval," IEEE Trans. Image Processing, 15(11), 3597-3601 (2006).
[20] Li, Y. and Shapiro, L.G., "Consistent line clusters for building recognition in CBIR," Proc. IEEE Int'l Conf. Pattern Recognition, Vol. 3, 952-956 (2002).
[21] Marcelja, S., "Mathematical Description of the Responses of Simple Cortical Cells," J. Optical Society of America, 70(11), 1,297-1,300 (1980).
[22] McLachlan, G.J., [Discriminant Analysis and Statistical Pattern Recognition], Wiley–Interscience, New York, (1992).
[23] Oliva, A. and Torralba, A., "Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope," Int'l J. Computer Vision, 42(3), 145–175 (2001).
[24] Roweis, S.T. and Saul, L.K., "Nonlinear Dimensionality Reduction by Locally Linear Embedding," Science, Vol. 290, 2,323-2,326 (2000).
[25] Rui, Y., Huang, T. S., Ortega, M. and Mehrotra, S., "Relevance Feedback: A Power Tool in Interactive Content-based Image Retrieval," IEEE Trans. Circuits and Systems for Video Technology, 8(5), 644-655 (1998).
[26] Siagian, C. and Itti, L., "Rapid Biologically-Inspired Scene Classification Using Features Shared with Visual Attention," IEEE Trans. Pattern Analysis and Machine Intelligence, 29(2), 300-312 (2007).
[27] Song, D. and Tao, D., "Biologically Inspired Feature Manifold for Scene Classification," IEEE Trans. Image Processing, 19(1), 174-184, (2010).
[28] Tao, D., Li, X., Wu, X. and Maybank, S.J., "General Tensor Discriminant Analysis and Gabor Features for Gait Recognition," IEEE Trans. Pattern Analysis and Machine Intelligence, 29(10), 1,700-1,715 (2007).
[29] Tao, D., Li, X., Wu, X. and Maybank, S.J., "Geometric Mean for Subspace Selection," IEEE Trans. Pattern Analysis and Machine Intelligence, 31(2), 260-274 (2009).
[30] Tao, D., Tang, X. and Li, X., "Which Components Are Important for Interactive Image Searching?," IEEE Trans. Circuits and Systems for Video Technology, 18(1), 1-11 (2008).
[31] Tao, D., Tang, X., Li, X. and Wu, X., "Asymmetric Bagging and Random Subspace for Support Vector Machines-based Relevance Feedback in Image Retrieval," IEEE Trans. Pattern Analysis and Machine Intelligence, 28(7), 1088-1099 (2006).
[32] Tenenbaum, J., Silva, V. and Langford, J., "A Global Geometric Framework for Nonlinear Dimensionality Reduction," Science, Vol. 290, 2,319-2,323 (2000).
[33] Tong, S. and Chang, E., "Support Vector Machine Active Learning for Image Retrieval," Proc. ACM Int'l Conf. Multimedia, 107-118 (2001).
[34] Treisman, A. and Gelade, G., "A Feature-integration Theory of Attention," Cognitive Psychology, Vol. 12, 97-137 (1980).
[35] Turner, R. S., [In the Eye's Mind: Vision and the Helmholtz Hering Controversy], Princeton University Press, (1994).
[36] Ullah, M.M., Pronobis, A., Caputo, B., Luo, J., Jensfelt, R. and Christensen, H.I., "Towards robust place recognition for robot localization," IEEE Int'l Conf. Robotics and Automation, 530-537 (2008).

[37] Vapnik, V., [The Nature of Statistical Learning Theory]. Springer-Verlag, (1995).
[38] Yan, S., Xu, D., Yang, Q., Zhang, L., Tang, X. and Zhang, H., "Multilinear Discriminant Analysis for Face Recognition," IEEE Trans. Image Processing, 16(1), 212-220 (2007).
[39] Zhang, T., Tao, D., Li, X. and Yang, J., "Patch Alignment for Dimensionality Reduction," IEEE Trans. Knowledge and Data Engineering, 21(9), 1299-1313 (2008).
[40] Zhang, W. and Kosecka, J., "Localization based on Building Recognition," IEEE Workshop on Applications for Visually Impaired, 21-28 (2005).
[41] http://eeepro.shef.ac.uk/building/dataset.rar
[42] http://www.ece.osu.edu/~maj/osu_svm/, 2002.