Vision and Information Engineering

Dimensionality Reduction-based Building Recognition

Jing Li and Nigel M. Allinson

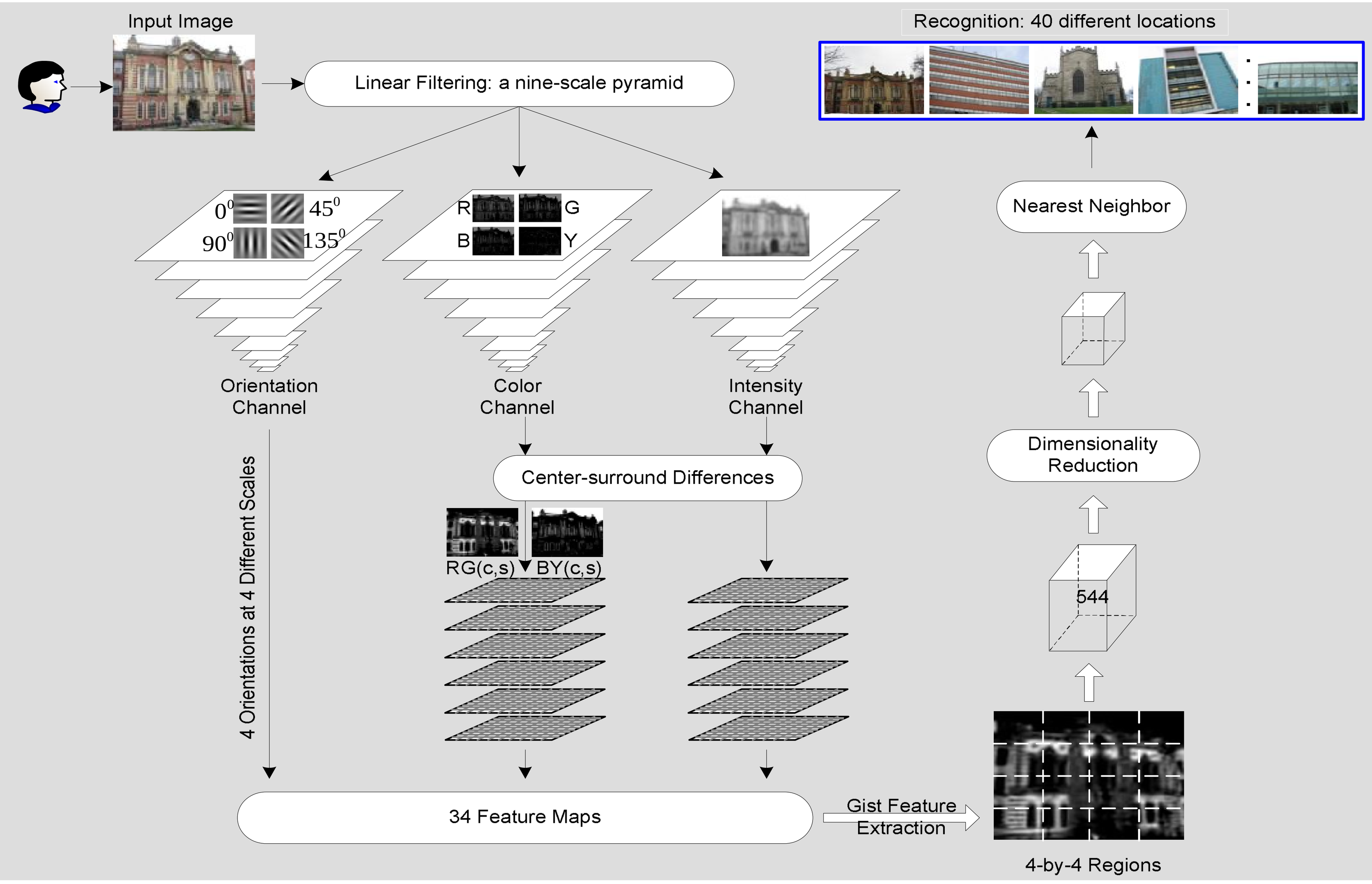
Building Recognition Scheme

Building Recognition aims to distinguish different buildings in a large-scale image database.

We proposed a building recognition scheme [2] based on biologically-inspired feature extraction and subspace learning-based dimensionality reduction.

Applications

i) architectural design
ii) building labelling in videos
iii) robot vision or localization
iv) mobile device navigation



The main stages in our scheme:
1) biologically-inspired feature representation [1];
2) dimensionality reduction by subspace learning algorithms; and
3) classification by nearest neighbour rule.

The advantages of our work:
1) extracted features are biologically related to the human visual perception;
2) features are invariant to scaling, rotation, illumination changes, and occlusions, and
3) each stage of our system requires low computational cost.

Motivations

Current building recognition algorithms achieves satisfactory results. However, they suffer from some limitations:
1) They are based on low level visual features, e.g., vanish point detection;
2) recognition is conducted on pairs of raw high-dimensional feature vectors, which results in high computational cost and memory requirements.

Feature Representation

- **Linear filtering**
  Each input image is linear filtered to give a Gaussian pyramid of nine scales $I(\vec{p};s)\quad s=0,\dots,8$
- **Visual feature extraction**
  1) Intensity and colour
  Centre-surround operation $I_{c,s}=|I_c \Theta I_s|$
  difference between the pixel at a centre scale $c=2,3,4$ and its corresponding pixel at a surround scale $s=c+d$ with $d=3,4$

  Intensity $I=(r+g+b)/3$

  Colour
  R, G, B, Y colour channel

$$R=r-(g+b)/2$$
$$G=g-(r+b)/2$$
$$B=b-(r+g)/2$$
$$Y=r+g-2(|r-g|+b)$$

  Three pairs of colour opponents, resulting in 18 feature maps.

$$RG_{c,s}=|(R_c-G_c)\Theta(R_s-G_s)|$$
$$BY_{c,s}=|(B_c-Y_c)\Theta(B_s-Y_s)|$$
$$I_{c,s}=|I_c\Theta I_s|$$

  2) Orientation
  Gabor filters with 4 different scales and 4 different orientations, making a total of 16 Gabor functions

  There are 34 feature maps in total.

- **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.

Dimensionality Reduction

i) Conventional linear dimensionality reduction techniques, e.g., principal component analysis (PCA) and linear discriminant analysis (LDA); and
ii) Manifold learning algorithms: locality preserving projections (LPP), supervised LPP (SLPP), semi-supervised discriminant analysis (SDA).

PCA

PCA is an unsupervised feature extraction algorithm, which projects data along the direction with the largest variance

$$C=(1/N)\sum_{i=1}^{N}(\vec{x}_i-\vec{m})(\vec{x}_i-\vec{m})^T$$
$$\vec{m}=(1/N)\sum_{i=1}^{N}\vec{x}_i$$

Solve the eigenvalue problem:

$$\lambda\vec{u}=C\vec{u}$$

The resulting subspace is spanned by

$$\vec{u}_1,\vec{u}_2,\dots,\vec{u}_L$$

associated with the largest eigenvalues.

LDA

LDA is a supervised learning algorithm. It separates samples from different classes far away while keeping those within the same class close to each other in the projected lower dimensional subspace.

$$U_{opt}=\arg\max_U\frac{U^TS_bU}{U^TS_wU}\qquad m_i=(1/N_i)\sum_{j=1}^{n_i}\vec{x}_{i;j}$$
$$S_b=(1/N)\sum_{i=1}^{c}N_i(\vec{m}_i-\vec{m})(\vec{m}_i-\vec{m})^T\qquad N=\sum_{i=1}^{c}N_i$$
$$S_w=(1/N)\sum_{i=1}^{c}\sum_{j=1}^{N_i}(\vec{x}_{i;j}-\vec{m}_i)(\vec{x}_{i;j}-\vec{m}_i)^T\qquad m=(1/N)\sum_{i=1}^{c}\sum_{j=1}^{N_i}\vec{x}_{i;j}$$

Solve the generalized eigenvalue problem

$$S_bU=\lambda S_wU$$

The resulting subspace is spanned by

$$U=\{\vec{u}_1,\vec{u}_2,\dots,\vec{u}_L\}$$

LPP/SLPP

LPP is a linear manifold learning algorithm, which is able to preserve the local neighbourhood structure for each data point by constructing an adjacency graph among the neighbours.

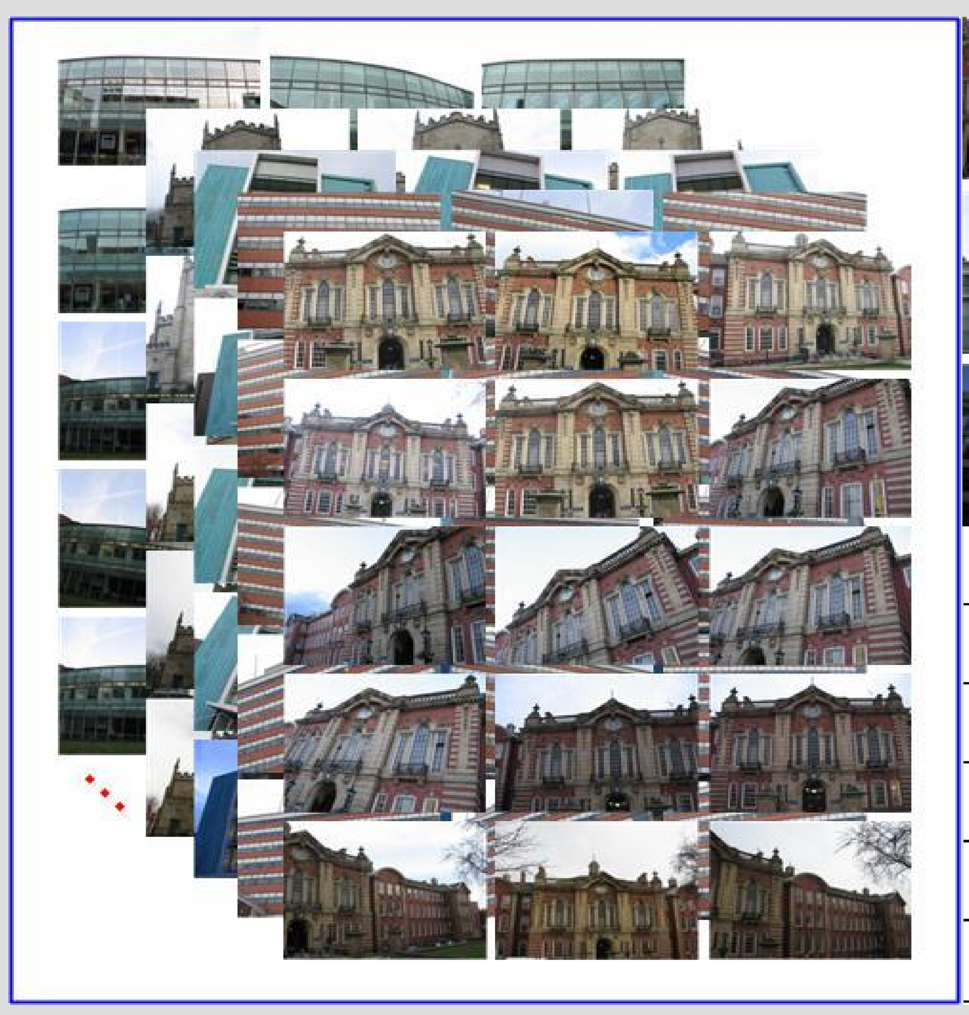
LPP is employed in an unsupervised mode, which utilizes -nearest neighbour search for graph construction; while SLPP is the supervised version of LPP by considering label information in adjacent graph construction.

SDA

SDA takes into account both labelled and unlabeled examples. Herein, labelled training samples are used to discriminate between different classes, whereas both labelled and unlabeled samples are applied to explore the manifold structure of the data points.

$$S_b=(1/N)\sum_{i=1}N_i\left((1/N_i)\sum_{j=1}\vec{x}_{i;j}\right)\left((1/N_i)\sum_{j=1}^{N_i}\vec{x}_{i;j}\right)^T=XWX^T$$
$$S_t=(1/N)\sum_{i=1}^{c}\sum_{j=1}^{N_i}(\vec{x}_{i;j})(\vec{x}_{i;j})^T=XX^T$$
$$U_{opt}=\arg\max_U\frac{U^TS_bU}{U^TS_tU}=\arg\max_U\frac{U^TXWX^TU}{U^TXX^TU}$$

Database and Performance Evaluation



	Recognition Rates (%)
PCA	47.36
LPP	51.40
SLPP	85.07
LDA	**85.25**
SDA	81.33

Our constructed database contains 3192 building images belonging to 40 categories. A variety of challenges include rotation, scaling, variant lighting conditions, viewpoint changes, occlusions, and vibration.

Results
1) LDA is the best, while PCA is the worst case;
2) SLPP performs much better than LPP; and
3) SDA, supposed to outperform LDA, does not obtain better results.

Conclusions:

1) Biologically-inspired features are especially robust to illumination changes;
2) Dimensionality reduction really works for building recognition; and
3) It is important to select an appropriate dimensionality reduction method for a specific application.

The database is available online: http://eeepro.shef.ac.uk/building/dataset.rar

References

[1] C. Siagian and L. Itti, Rapid Biologically-Inspired Scene Classification Using Features Shared with Visual Attention, IEEE Trans. Pattern Analysis and Machine Intelligence, 29(2) (2007) 300-312.
[2] Jing Li and Nigel M. Allinson, 'Dimensionality Reduction in Building Recognition', submitted.

The University Of Sheffield.

Further details, please contact:
Professor Nigel M Allinson
Dept. of Electronic and Electrical Engineering
Mappin Street, University of Sheffield
Sheffield, S1 3JD, UK
Phone: (+44)(0) 114 222 5412
Email: n.allinson@sheffield.ac.uk