TELKOMNIKA, Vol.17, No.1, February 2019, pp.473~480
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v17i1.11609

A colour-based building recognition using support vector machine

Mas Rina Mustaffa*[1], Loh Weng Yee[2], Lili Nurliyana Abdullah[3], Nurul Amelina Nasharuddin[4]
Multimedia Department, Faculty of Computer Science and Information Technology,
Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor Darul Ehsan, Malaysia
*Corresponding author, email: MasRina@upm.edu.my[1], lohwengyee94@gmail.com[2],
liyana@upm.edu.my[3], nurulamelina@upm.edu.my[4]

Abstract

Many applications apply the concept of image recognition to help human in recognising objects simply by just using digital images. A content-based building recognition system could solve the problem of using just text as search input. In this paper, a building recognition system using colour histogram is proposed for recognising buildings in Ipoh city, Perak, Malaysia. The colour features of each building image will be extracted. A feature vector combining the mean, standard deviation, variance, skewness and kurtosis of gray level will be formed to represent each building image. These feature values are later used to train the system using supervised learning algorithm, which is Support Vector Machine (SVM). Lastly, the accuracy of the recognition system is evaluated using 10-fold cross validation. The evaluation results show that the building recognition system is well trained and able to effectively recognise the building images with low misclassification rate.

Keywords: *10-fold cross validation, building, content-based, colour feature, support vector machine*

1. Introduction

Image recognition is a process to identify and detect the features in a digital image or video [1]. Many recognition applications are created such as for face [2, 3], voice [4], handwriting [5], agriculture [6-8], and others [9-11].

In an image recognition system, the algorithm is trained to be able to analyse and identify a given image. By training the algorithm, it will be able to perform visual tasks on its own. During the training phase, we need to tell the algorithm what to learn and how to do it. Firstly, each image in the database will be given a specific label. Then, the algorithm will be taught on how to compare the input image with the images in the database and classify them into the correct category.

Different types of domain uses different feature extraction methods. Every domain has its own special features. In this paper, the focused domain is building. A suitable feature extraction method is very important as the obtained feature values will be used later to train the system for recognising the building images. Therefore, few related previous works are studied in order to investigate the current feature extraction and recognition approaches used.

Authors in [12] allow users to upload a building image to the system and the system will analyse and recognise the given building image. The dataset used in this system are 50 building images taken from Ulsan Metropolitan City in Korea. In this system, it uses the combined features of edge detection, colour and keypoint detection to analyse and recognise a given building image. Firstly, canny edge detector is used to detect the line segments of buildings' facet. Segments of non-building pattern will be rejected. Next, colour histogram is used to detect the colour of the wall. Firstly, the building image is converted to gray scale image. All the pixels of the wall are extracted and a hue histogram with 36 bins is calculated. The highest peak in the histogram is detected and clustered into separate groups. The system further refines the recognition by extracting keypoints from the building image using Scale-Invariant Feature Transform (SIFT) descriptor. The advantages of this system are the system is invariant to image scale, rotation, noise, change in illumination and also occlusion. This system can still recognise buildings with small or large scale. Buildings that are rotated can be successfully recognised by

Received May 30, 2018; Revised October 9, 2018; Accepted November 12, 2018

the system. This building recognition system can deal with problems such as noisy images and having too high or too low light intensity. However, the system has some limitations such as the system has a complex feature extraction process to represent an image. It means that the feature representation requires more computational time and cost.

The system in [13] utilised a simpler way to recognise a building. There are two modules in this system which is segmentation of building facades region module and facade recognition module. The dataset used in this system are 1005 images of Zurich Building Image Database (ZuBuD) and images taken at Campus of University of Gavle. In the first stage, colour feature is being used. The system will segment out sky and other blue components by setting a threshold for blue value to find the locations. The rest will be set to black colour. The remaining sky image is transformed to HSV domain and mean values of the H channel is calculated. In the recognition stage, the building is recognised by identifying the vertical edges of the building. Chamfer distance value has been used to calculate the distance between the query images and the database images. This system provides a satisfactory recognition rate and it is invariant to rotation and image scale. Buildings in different angles and different scales can be successfully recognised by the system. However, the system might fail to recognise when the building images do not have similar lighting conditions.

The system proposed by Kaya and Kayci [14] allow users to upload the butterfly image to the system and the system will help to recognise and classify the species of the given butterfly image. This system uses 140 of butterfly images for the purpose of identifying 14 different butterfly species. This system uses the combination of colour and texture features to analyse and classify the butterfly species. Texture features such as Gray Level Co-occurrence Matrix (GLCM) with different angles and distances have been used in this system to measure the statistical characteristics of an image. The five GLCM features such as homogeneity, contrast, energy, correlation and entropy are combined with another three colour features which are the mean gray level of the Red, Green and Blue. These eight features will be combined and form a feature vector. These feature values will be used to identify and classify the butterfly species. The system has achieved a recognition rate of 92.85% by dividing the dataset into training-test sets of 70-30%. It proves that the combination of colour and texture features can be successfully used in recognising different butterfly species. However, the system might not be able to recognise a butterfly sample which is deteriorated or discoloured.

Based on the few existing works on recognition, it has been found that different types of recognition system use different feature extraction methods. For example, system [9] uses SIFT descriptor to extract the special features for each building. Besides, features such as colour and texture have been used in system [13] and [14] to analyse and recognise building images and butterfly images. System in [14] uses the features of colour and texture. Colour has been proven by [13] that it can be used to recognise a given building image. Also, each building has its own unique building facade texture. Therefore, the building facade texture for each building can be taken into consideration as one of the features to represent a building image.

Ipoh is the capital city of the state of Perak in Malaysia. In the previous years, Ipoh might be just a small city that has many old buildings and do not gain much attention from others. Nowadays, the city is slowly developing where more modern buildings can be seen around the city. The old buildings in Ipoh eventually become the historical buildings. Recently, Ipoh has been ranked as one of the top spots in Asia by the largest travel guide book publisher in the world, Lonely Planet [15]. All of the interesting parts in Ipoh are starting to gain attention from people all over the world.

In order to allow people to search and know more about the buildings in Ipoh in a more fast and accurate way, the concept of image recognition plays an important role. People have been developing building recognition systems using different feature extraction methods and each building recognition system will focus on their preferred area. Based on the research on existing building recognition systems, it has been found that there is no building recognition system that focuses on Ipoh area has been created. Ipoh is a city that needs a continuous development. Therefore, this work is about proposing a recognition framework to recognise and classify a given building image. We want to develop something which can help more people to know more about Ipoh. This prototype building recognition system allows user to upload a building image to the system. Then, the system will recognise and classify the building image. The outline of this paper is as follows. Section 2 explains the proposed method while section 3 describes the building recognition system. The framework for evaluation and analysis of results on the other hand are described in section 4. Finally, the conclusion and future direction are presented in section 5.

2. Research Method

The recognition process have two phases which are training phase and recognition phase. In training phase, SVM model will be used to train the system to recognise the four categories of buildings using the obtained feature values. Then, the system will predict the building categ ory of an input image during the recognition phase. The overall framework of a recognition system is SHOWN in Figure 1.

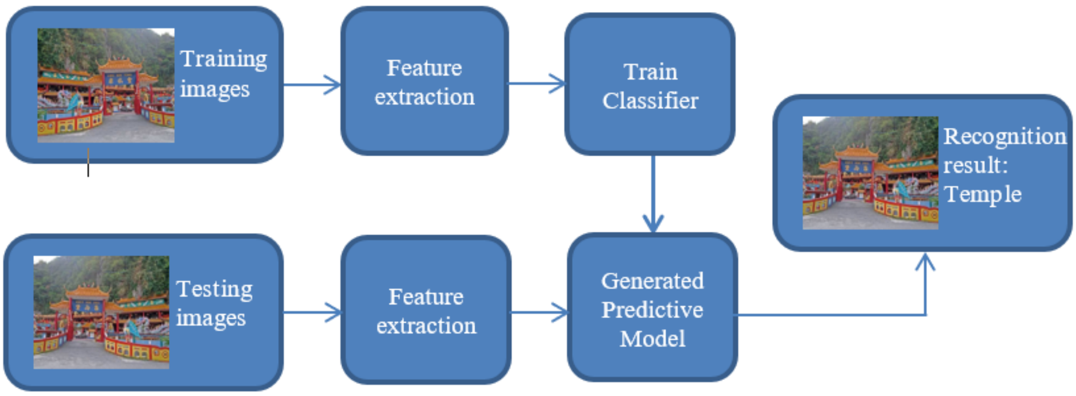

Figure 1. Complete framework of the building recognition system

In feature extraction process, the input image will first be converted into a gray scale image as gray scale consists only 256 levels. A gray scale image only need one byte for each pixel. Therefore, it is less complex when compared to RGB colour space which has a total of three bytes for each pixel. Since the size of the each image might vary, it will be resized to a standard dimension. Based on the experimental result, it has been found out that 128×128 is the most suitable image size that gives the best result. Colour histogram is then calculated where it considers the distribution of colours in an image. It is a method that is invariant to rotation and translation. It provides a statistical graph that shows the number of pixels that appeared in different types of colour. Then, a 3×3 laplacian mask is applied to sharpen the input image. Laplacian mask is a second order derivative mask. It is better as compared to other first order derivative masks because it will undergo two derivative processes. A 3×3 laplacian filter mask is chosen as 3×3 mask size is commonly used and not too complex. Using the probability distribution of each intensity levels, the mean (μ), standard deviation (σ), variance (σ^2), skewness (θ), and kurtosis (γ) are computed to form the feature vector based on the respective (1-5) below.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (1)$$

$$\sigma = \sqrt{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2\right)} \quad (2)$$

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2 \quad (3)$$

$$\theta = \sqrt[3]{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^3\right)} \quad (4)$$

$$\gamma = \sqrt[4]{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^4\right)} \quad (5)$$

where N is the total number of pixels within an image and x_i is the current pixel being processed. The statistical features provided in the histogram shows more about the characteristics of the intensity levels in an image. For example, a low contrast image will have a narrow histogram while a brighter image will have a wider histogram. Therefore, in total there are five features used as the representation for each image.

The system will then be trained to recognise building images by using Support Vector Machine (SVM). Each train image will be associated with a class label. Before deciding on SVM, we have conducted a mini experiment by training the training images using various classifiers such as SVM, Tree, Linear Discriminant, *K*-Nearest Neighbour, and few others. From the empirical studies, we have found that SVM performs the best for our work. Support Vector Machine (SVM) is a type of supervised learning model that is used for classification task. SVM is usually used to classify one or two classes. Therefore, if three or more classes need to be classified, a normal SVM could not be used but instead a multiclass SVM is a good approach. It uses a one-to-one coding design where it will reduce the classification problem to a set of binary classifier. For example, if there are three classes, using the formula $K(K-1)/2$ where K=number of classes, a total of three binary classifier will be created. The three binary classifiers are used to train in differentiating between classes at a time. To test a new input, all the three binary classifiers will first be applied to the new input. Hamming distance, as shown in (6) is used as the distance measure and the new input will be classified to the class with the nearest distance.

$$d^{HAD}(i,j) = \sum_{k=0}^{n=1}\left[y_{i,k} \neq y_{j,k}\right] \tag{6}$$

where d^{HAD} is the Hamming distance between the objects i and j, k is the index of the respective variable reading y out of the total number of variables n. The Hamming distance itself gives the number of mismatches between the variables paired by k.

Figure 2 (a) shows the process flow of the proposed colour-based feature extraction while Figure 2 (b) visualises the SVM algorithm process.

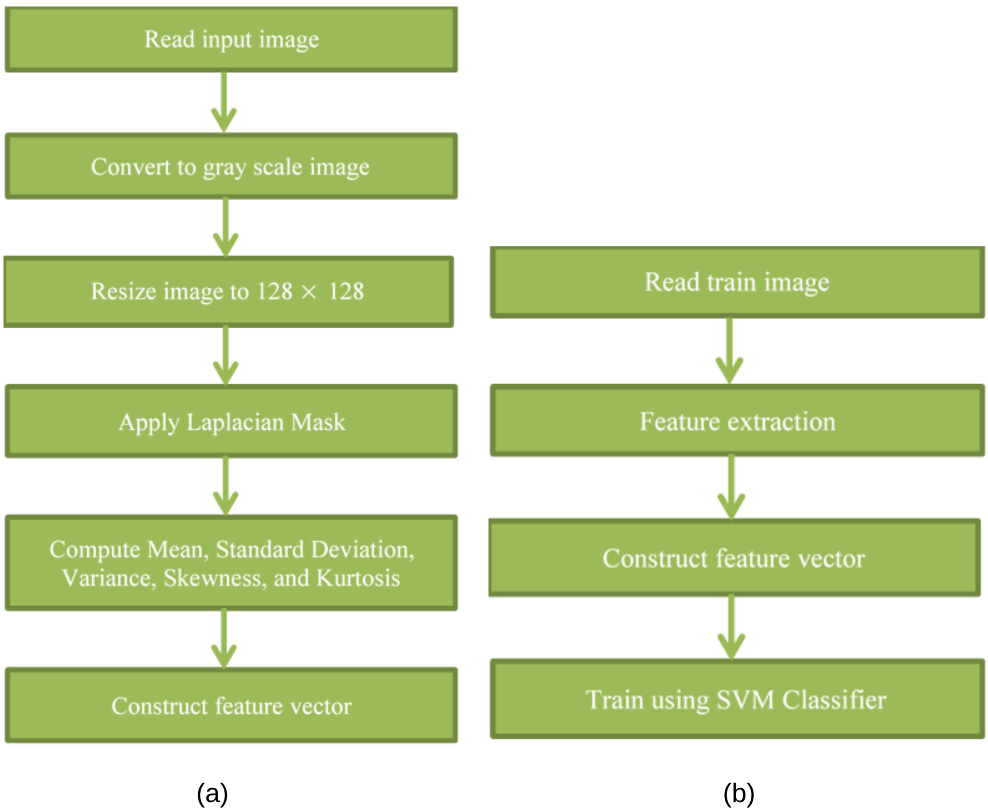

Figure 2. Process flow for (a) colour-based feature extraction
(b) training of images based on SVM algorithm

3. System Specification

The proposed building recognition framework is developed on Intel®Core™i5-4210U 1.7GHz with 64-bit Windows 8.1 operating system, and 4GB Random Access Memory. The software requirements are MATLAB R2016a, MySQL Workbench 6.3 CE and MySQL Server 5.7. MATLAB is used to develop the proposed system, MySQL Workbench is used to create all the tables and MySQL server is used to connect the database to MATLAB. Figure 3 shows the Entity–Relationship diagram of the proposed building recognition system. Table building_image is the main table where the file_name of each building image will act as a primary key. File_name also acts as a foreign key in table histogramfeature.

Figure 4 shows an interface where there is a list box showing all the training images that have already been saved into the database and will be used later for recognition process. In addition, there are few function buttons such as upload, update, and delete button in this interface. Data can be uploaded to the database, details that already saved into the database can be updated or data can be deleted from the database. Figure 5 shows the interface where a building image can be uploaded to the system and the result predicted by the system will be displayed in a text box.

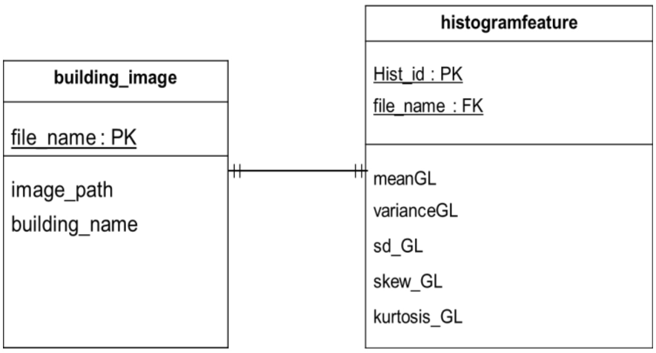

Figure 3. Entity–Relationship diagram for building recognition system

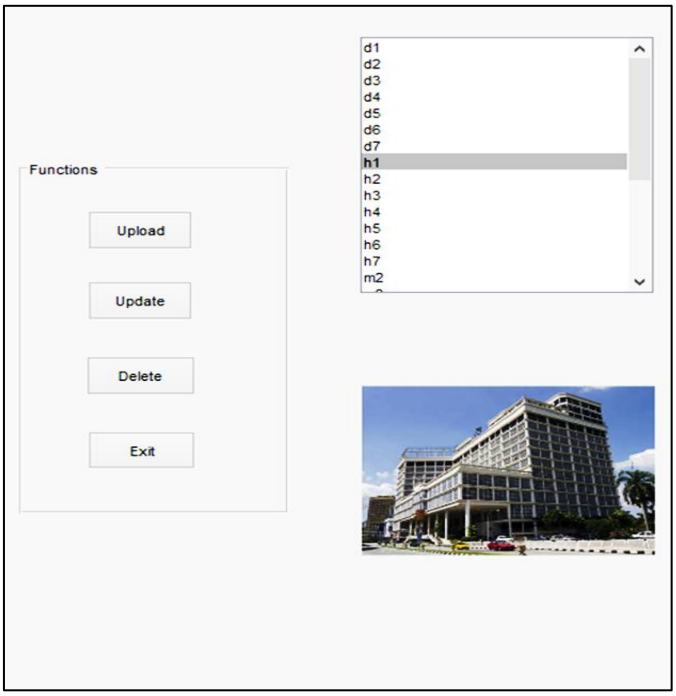

Figure 4. Graphical user interface for upload training image interface

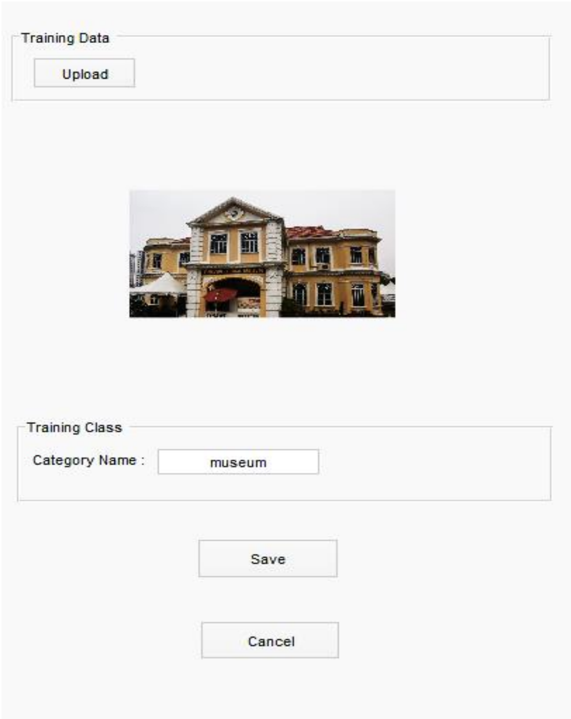

Figure 5. Graphical user interface for building image recognition

4. Results and Discussion

There is a total of four categories of buildings which are temple, museum, hotel and mosque. Each category consists of 10 images which result in 40 images. All the building images are obtained from the Internet. These categories are utilised since many buildings found in Ipoh, Perak, Malaysia are of these types. Samples of the dataset are shown in Figure 6.

Figure 6. Samples of building images consist of museum, hotel, mosque, and temple

10-fold cross validation will be used to evaluate the accuracy of the recognition system. It is a technique used to evaluate the system by randomly partitioning the data into 10 equal size of subsamples. Nine subsamples will be used to train the model and one subsample will be used as testing data. This cross-validation process will be repeated for 10 times. Then, the accuracy of the overall system will be obtained by calculating the average of the 10 generated results. The proposed feature extraction method will be compared with the feature extraction method used in [14]. Table 1 shows the summary of the individual accuracy value for 10 folders for both approaches.

Table 1. Summary of Individual Accuracy Value of Cross-validation for 10 Folders of Our Proposed Work and Method by [14]

Folder	Accuracy of cross-validation (%) Proposed method	Method by [14]
1	100	75
2	100	75
3	100	75
4	100	100
5	100	75
6	75	75
7	100	75
8	100	100
9	75	100
10	100	75
Average	95	82.5

Based on Table 1, it can be seen that the recognition system using the proposed feature extraction method has able to achieve an overall higher accuracy in comparison to the recognition system using the benchmark feature extraction method. This result shows that the benchmark feature extraction method may not be so effective when it is used on Ipoh building images. The percentage of the correct classification from each category sums up to an overall accuracy of 82.5% for the benchmark method. 17.5% of misclassification rate is obtained by the benchmark method which shows that the feature extraction method is not as effective as compared to the recognition system using the proposed feature extraction method. The proposed framework on the other hand is able to obtain 95% recognition accuracy with only 5% misclassification.

5. Conclusion and Future Research

This paper proposed a building recognition framework which is able to recognise various buildings. Colour histogram has been used as the feature extraction method to extract the special feature of each building image. After obtaining the feature values of the building images, the algorithm is trained to analyse and recognise building images using Support Vector Machine (SVM). Based on the conducted experiments, it is shown that the proposed building recognition system is well-trained to effectively recognise building images in Ipoh city, Perak, Malaysia, and has achieved an overall accuracy of 95%.

For future work, we plan to expand the dataset by including more images for each building category as well as exploring other building categories. A mobile-based application will also be considered to make it easier for general users or tourists to utilise the app when touring around Ipoh city.

Acknowledgement

This work is supported by Universiti Putra Malaysia under the Putra Grant project code GP/2018/9644500.

References

[1] Sonka M, Hlavac V, Boyle R. Image processing, analysis, and machine vision. Fourth Edition. Standford, USA: Cencage Learning. 2015.

[2] Ding C, Choi J, Tao D, Davis L. Multi-directional multi-level dual-cross patterns for robust face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2016; 38(3): 518-531.
[3] Ding C, Tao D. Robust face recognition via multimodal deep face representation. *IEEE Transactions on Multimedia*. 2015; 17(11): 2049-2058.
[4] Boles A, Rad P. Voice biometrics: *Deep learning-based voiceprint authentication system*. Proceedings of the 2017 12th Conference on System of Systems Engineering (SoSE). Waikoloa, Hawaii, USA. 18-21 June 2017: 1-6.
[5] Kumar P, Saini R, Roy PP, Pal U. A lexicon-free approach for 3D handwriting recognition using classifier combination. *Pattern Recognition* Letters. 2018; 103: 1-7.
[6] Nasirahmadi A, Ashtiani SHM. Bag-of-Feature model for sweet and bitter almond classification. *Biosystems Engineering*. 2017; 156: 51-60.
[7] Xu, L, Lv, J. Recognition method for apple fruit based on SUSAN and PCNN. *Multimedia Tools and Applications.* 2018; 77(6): 7205-7219.
[8] Pereira, LFS, Barbon, S, Valous, NA, Barbin, DF. Predicting the ripening of papaya fruit with digital imaging and random forests. *Computers and Electronics in Agriculture.* 2018; 145: 76-82.
[9] Kahou, Samira Ebrahimi, Xavier Bouthillier, Pascal Lamblin, Caglar Gulcehre, Vincent Michalski, Kishore Konda, Sébastien Jean et al. Emonets: Multimodal deep learning approaches for emotion recognition in video. *Journal on Multimodal User* Interfaces. 2016; 10(2): 99-111.
[10] Escalera S, Athitsos V, Guyon I. Challenges in multi-modal gesture recognition. *Gesture Recognition*. 2016: 1-60.
[11] Llano EG, Vázquez MSG, Vargas JMC, Fuentes LMZ, Acosta AAR. Optimized robust multi-sensor scheme for simultaneous video and image iris recognition. *Pattern Recognition Letters.* 2018; 101: 44-51.
[12] Trinh H, Kim D, Jo K. Supervised training database for building recognition by using cross ratio invariance and SVD-based method. *Applied Intelligence.* 2010; 32(2): 216-230.
[13] Ahlen J, Seipel S. *Knowledge Based Single Building Extraction and Recognition*. *Recent Advances in Computer Engineering, Communications and Information* Technology. 2014: 29-35.
[14] Kaya Y, Kayci L. Application of artificial neural network for automatic detection of butterfly species using color and texture features. *The Visual Computer.* 2014; 30(1): 71-79.
[15] Planet, L. (n.d.). Ipoh. Retrieved April 08, 2018, from https://www.lonelyplanet.com/malaysia/peninsular-malaysia-west-coast/ipoh