

Chapter 2

Transformation Invariant Image Recognition Using Multilayer Perceptron

2.1 Introduction

A multilayer perceptron (MLP) [52, 53] comprises an input layer, any number of hidden layers and an output layer, enjoying full interconnectivity among themselves. As mentioned in Chap. 1, an MLP architecture is efficient in classifying nonlinearly-separable data points due to the presence of the hidden layers in between the input and output layers of the network architecture.

This chapter illustrates two applications of the supervised MLP architecture [52, 53] for the recognition and detection of true color images. The first of these applications is centered around training the multilayer perceptron [52, 53] with a series of images, followed by the recognition/retrieval of matching images. The second application focusses on the efficiency of the multilayer perceptron [52, 53] in detecting different transformations viz., scaling, translation and rotation in test images after the network architecture is properly trained with non-transformed versions of the test images.

Section 2.2 illustrates the proposed methodology adopted in training the multilayer perceptron [52, 53] with images and further recognition/retrieval of matching images when a subset of the training images are presented to the trained network. Section 2.3 elaborates on the findings of the image recognition/retrieval process by the MLP architecture [52, 53]. Section 2.4 discusses the algorithm and application of the multilayer perceptron [52, 53] for detecting different transformations (viz., scaling, translation and rotation) in the test images, once the network has been trained with non-transformed versions of these images. Section 2.5 sums up the results of detection of different transformations by the MLP network architecture. Section 2.6 draws a line of conclusion to the chapter.

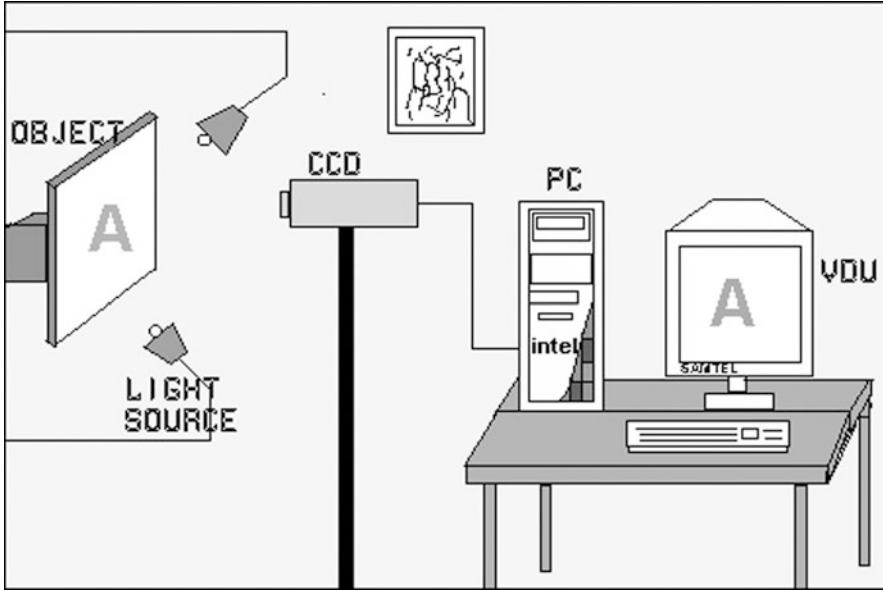


Fig. 2.1 Experimental arrangement

2.2 Image Recognition by an MLP Architecture

The procedure of image recognition followed by the detection and evaluation of transformations in test/target images has been carried out by means of a trained MLP architecture [52, 53]. A multilayer perceptron [52, 53] comprising eight input nodes, eight hidden nodes and eight output nodes is employed for training a series of images followed by the recognition task.

The experimental arrangement employed for the implementation of the image recognition and detection task is shown in Fig. 2.1. The arrangement comprises the following hardware components.

1. A workstation/PC
2. A frame grabber card built into the workstation/PC
3. A CCD camera for image acquisition
4. Light sources to illuminate the image scene

The flowchart of the recognition and transformation detection procedure is shown in Fig. 2.2. The different steps of the image recognition/transformation detection process by a trained MLP architecture [52, 53] are described in the following subsections.

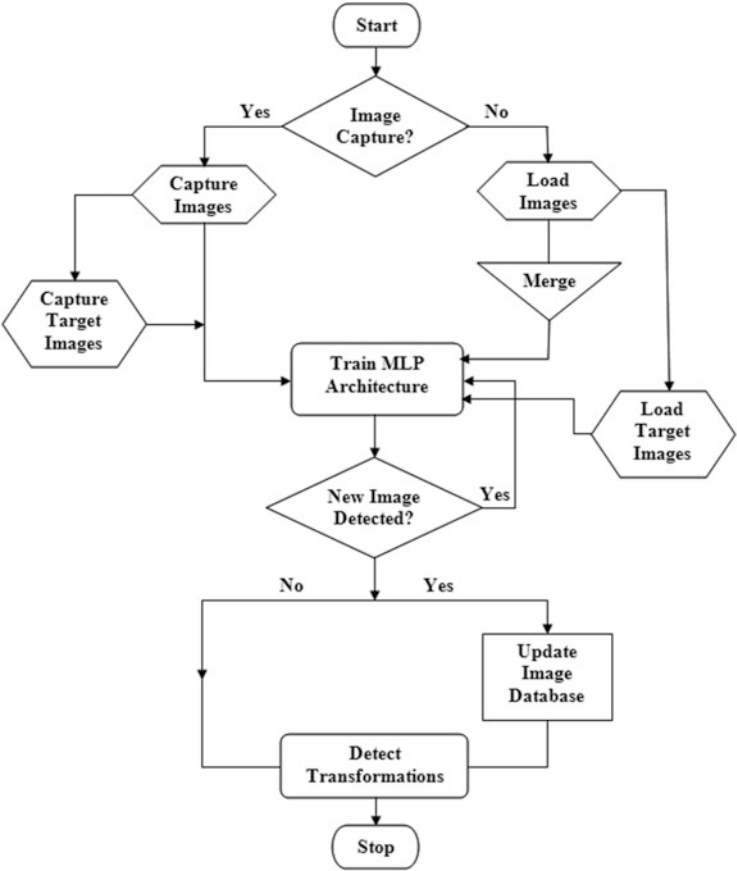


Fig. 2.2 Flowchart of the transformation invariant image recognition and detection process

2.2.1 Image Acquisition Phase

This is the starting and the most vital phase of the application. It acquires both standard test images and real-life images to train the inherent MLP architecture [52, 53]. It may be noted that the application presented here also bears a provision of recognizing real-life images from the external world. For the purpose of acquisition of real-life images, a CCD camera is employed.

As stated earlier, the image recognition task starts with either acquiring real-life images from the external world or loading images from an image database for training the MLP architecture [52, 53] based on the user’s choice. If images are captured by the CCD camera, test/target images are also captured likewise and stored in the database. Similarly, artificial test/target images are also stored in the database. These test/target images are used up in the recognition phase.

The application also generates new images by merging images stored in the database. This is evident from the flowchart.

2.2.2 Training the MLP Architecture

The MLP architecture [52, 53] is trained using the images stored in the database by means of the standard backpropagation algorithm. These images comprise either standard test images or real-life images acquired by means of a CCD camera from the external world. Once the MLP [52, 53] is trained, it is used faithfully for recognizing the test/target artificial/real-life images.

2.2.3 Recognition of Test Images

This is the third phase of the application, where the recognition/retrieval of images pertaining to input test images is carried out. Either standard test images or real-life images are presented to the trained MLP architecture [52, 53], whereby the MLP architecture [52, 53] recalls the image which closely resembles the presented image. It may be noted at this juncture that MLP [52, 53] has a high-level of approximation capabilities; hence even when trained images corrupted with noise or parts of trained images are presented to the network architecture, MLP can recall the correct image successfully. If, however, a completely new test/target image different from any of the training set of images is presented to the MLP [52, 53], it is subsequently stored in the database and the database is updated accordingly. Moreover, the application allows the trained MLP architecture [52, 53] to train on the newly recognized test/target image, as is evident from the flowchart.

2.2.4 Detection of Transformations

After the recognition phase is over, detection of transformations in the test/target images is carried out. The recognized images are sent to the detection module (to be described later) for the purpose of detection of different transformations inherent in them. This may be scaling, translation or rotation.

2.3 Results of Image Recognition

The multilayer perceptron- (MLP-) [52, 53] based image recognition task is carried out on several artificial images shown in Fig. 2.3. These include Image#1 (Lena image), Image#2 (Mandrill image) and Image#3 (Peppers image) each of dimensions

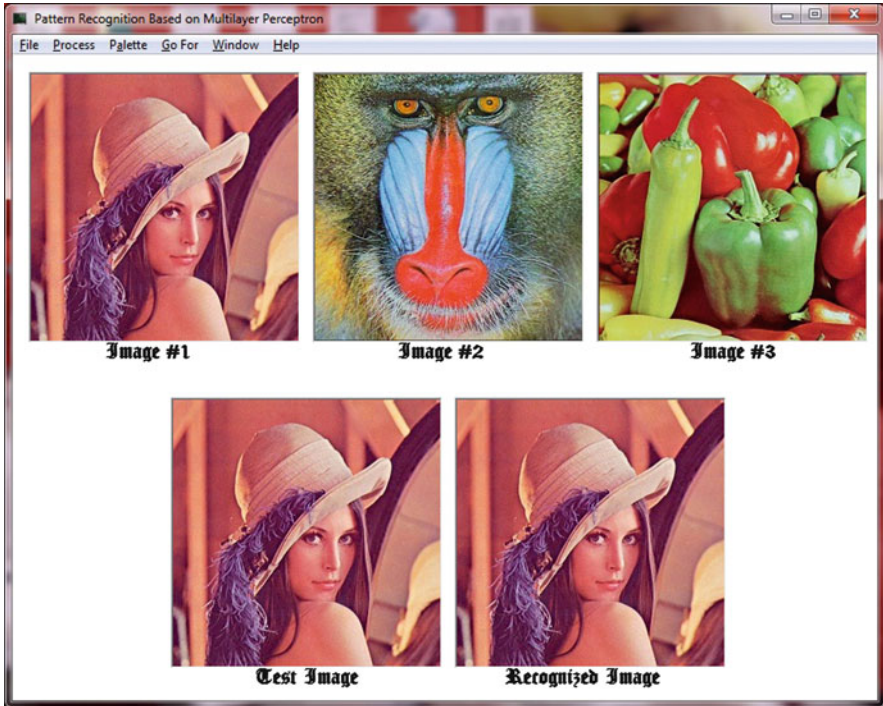


Fig. 2.3 Recognition of Lena image

of 256×256 . These images are used to train the MLP architecture using the standard backpropagation algorithm for the weight adjustment and error compensation technique. Figures 2.3–2.5 show the results of recognition when the test images belong to a subset of the training set of images. Figures 2.6–2.8 show the generalization capabilities of the MLP [52, 53] trained architecture.

Figure 2.6 shows that the network is able to recognize faithfully the underlying image even if an image from the training set masked by a huge amount of noise is used as the test image. This shows the noise immunity of the MLP architecture [52, 53]. Moreover, it is evident from Fig. 2.7 that once the MLP is precisely trained, it can recognize an image even if only a small part of the entire image is presented to it. This also reiterates the generalization capabilities of the MLP architecture.

Another interesting aspect of the MLP [52, 53] behavior is brought to light when a merged image formed by equal number of pixels from two images is used as a test image. Figure 2.8 shows such a situation. The test image is an admixture of the information of Image#1 and Image#2. It is seen from the results that the MLP [52, 53] recalls Image#1 of its training set in these circumstances. This can be attributed to the fact that in the event of a tie (which is the case with the test image comprising equal information from Image#1 and Image#2), the image from the training set which is first used for training the network gets recalled.

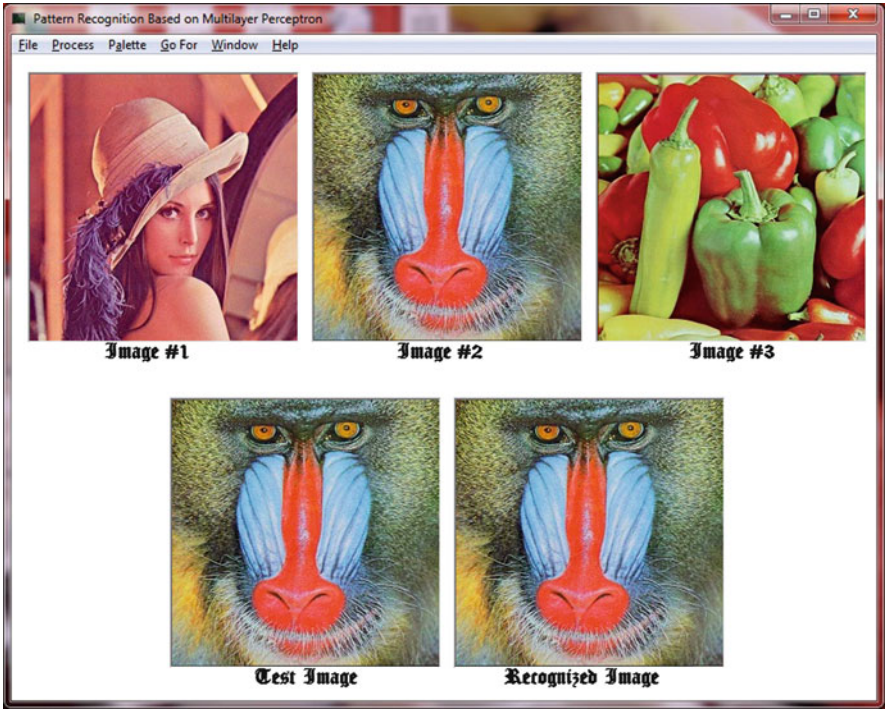


Fig. 2.4 Recognition of Mandril image

2.3.1 Results of Real-Life Image Recognition

Finally, a real-time application of the MLP architecture [52, 53] for recognition of real-life images is presented. Figure 2.9 shows an example of the recognition of real-life images acquired from the external world. Here, three real-life images are acquired by a CCD camera to form the training set of images. Once the network is trained with these images, a new real-life image (which, incidentally, is a transformed version of one of the training set images) captured from the external world is used as a test image. When this real-life test image is presented to the network, the network faithfully recognizes it and recalls the corresponding image from the training set. Thus, it is evident from the results of recognition of true color artificial images and true color real-life images acquired from the external world that a properly trained MLP architecture [52, 53] is able to recognize faithfully both types of images.

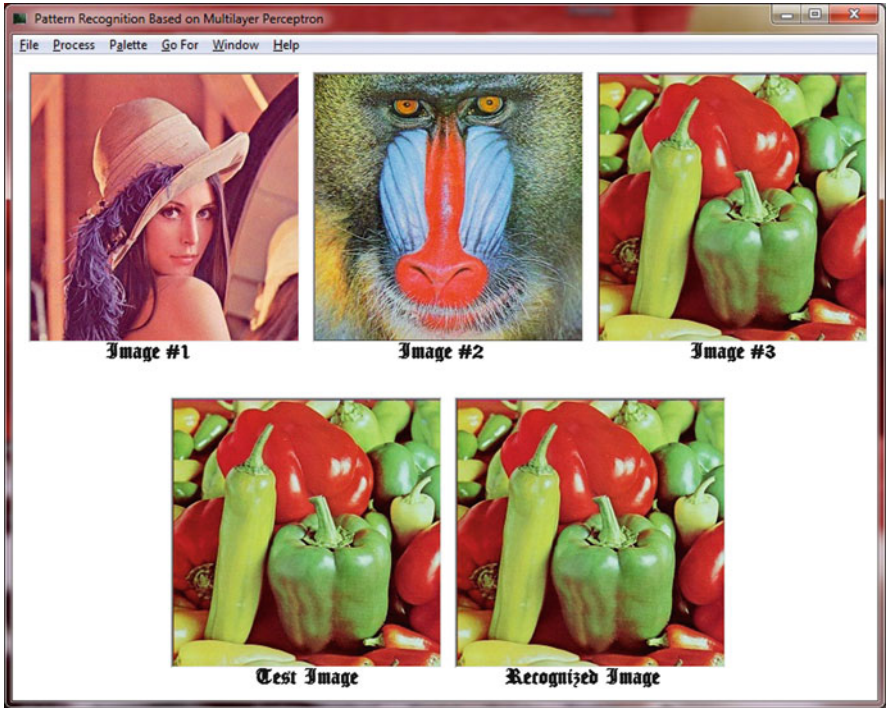


Fig. 2.5 Recognition of Peppers image

2.4 Transformation Detection by an MLP Architecture

As mentioned in Sects. 2.2 and 2.3, a multilayer perceptron [52, 53] is efficient in recognizing images from non-noisy, noisy and degraded versions of trained images. The training paradigm of the MLP architecture [52, 53] aims at obtaining processed cumulative information of the input image pixels at the output nodes. When the training of the MLP architecture [52, 53] gets completed, the individual input nodes are tied to the hidden nodes by some distinct interconnection weights. Similarly, distinct interconnection relationships also exist between hidden and output nodes after the training process. These interconnection relationships between either the input-hidden or hidden-output nodes of the MLP architecture [52, 53] can be put to use to detect transformations in test images.

This section illustrates the application of these interconnection relationships for the detection and evaluation of various image transformations like scaling, translation and rotation.

A multilayer perceptron [52, 53] model with eight input nodes, eight hidden nodes and eight output nodes is used to train a series of images. The input-hidden

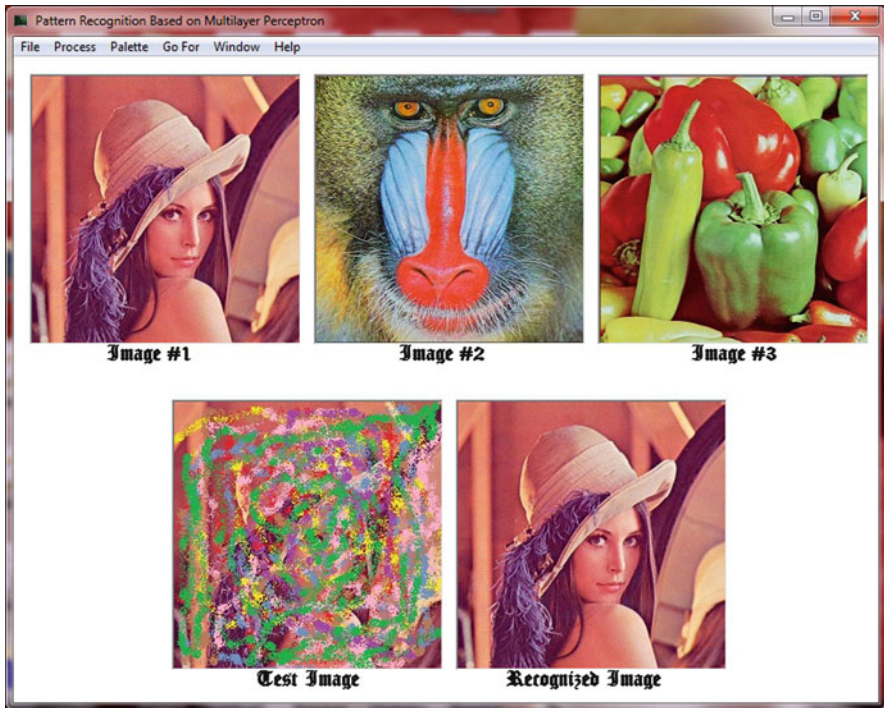


Fig. 2.6 Recognition of Lena image from noisy version

layer and hidden-output layer interconnection weights are initially set to zero. The network is then trained using a learning rule of the form:

$$w_{ijkl} = (d_i - y_i)x_j x_k x_l \quad (2.1)$$

where d_i and y_i are the training and actual outputs, respectively, for the i th image. x_j , x_k , x_l are the three inputs from three points in the image.

After the training is over, all the possible sets of inputs, which when joined would form a triangle, get connected similarly to their hidden nodes.

When the test images are presented to the network, these distinct input-hidden node interconnection relationships are applied on the individual pixels of the test images. Given these input-hidden node interconnection relationships, only those inputs from the test images which are similar to the images used up for training would yield similar responses at the corresponding hidden nodes. Thus, a subset of the test image pixels would form triangles which are similar to those formed in the original images, when these input-hidden node interconnection relationships are applied on them.

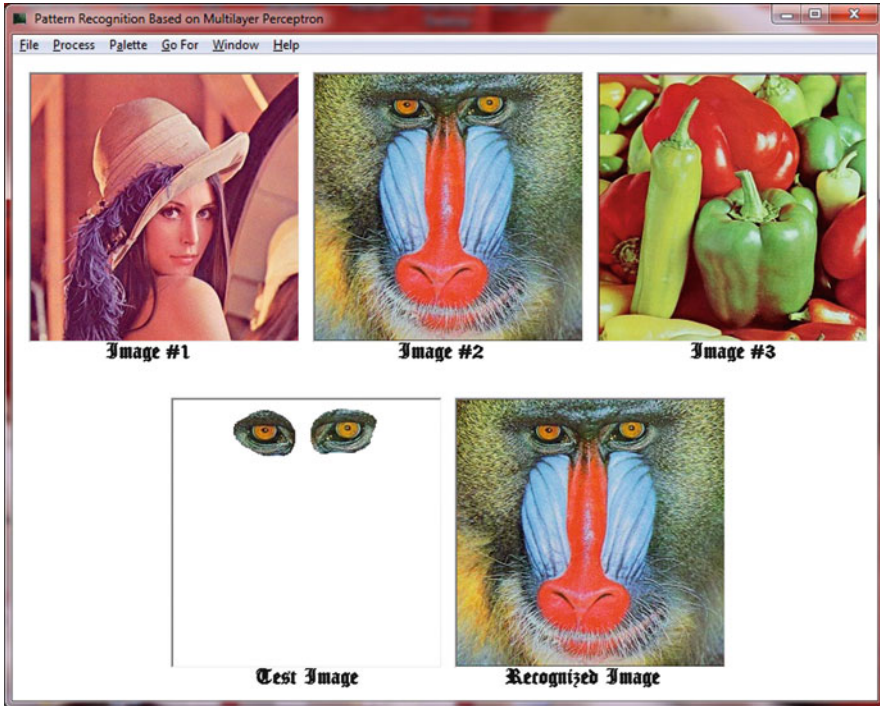


Fig. 2.7 Recognition of Mandril image from a part of the image

Once these sets of three inputs are identified in both the training and test images, comparison of the different attributes of the similar triangles formed by these inputs would yield the requisite detection and evaluation of the different image transformations as follows:

- A comparison of the lengths of the largest sides of the similar triangles formed in the original and test images would give the amount of scaling in the test image.
- Similar comparison along with the comparison of the starting coordinates of the largest sides of the similar triangles would give the amount of translation in the test image.
- A comparison of the slopes of the largest sides of the similar triangles formed in the original and test images would give the amount of rotation in the test image.
- If the starting coordinates of the largest sides of the similar triangles formed in the original and test images differ from each other, this means that the test image has undergone translation. Otherwise, there is no translation in the test image.
- If the lengths of the largest sides of the similar triangles formed in the original and test images are same, it implies that the test image is of the same dimensions as the original image. Otherwise, the test image is scaled.

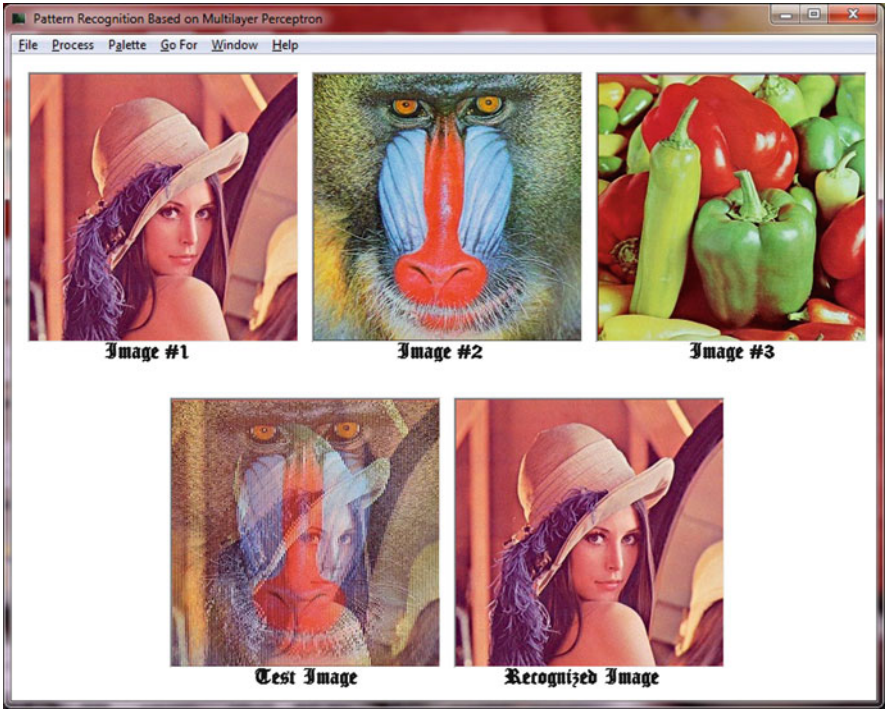


Fig. 2.8 Recognition of Lena image from merged image

- If the slopes of the largest sides of the similar triangles formed in the original and test images are same, it implies that the test image bears the same orientation as the original image. Otherwise, the test image is rotated with respect to its original counterpart.

2.5 Results of Transformation Detection

The detection of different image transformations by an MLP architecture [52,53] is demonstrated on the Lena and Plus images (Figs. 2.10–2.16). Several versions of the images, with the three different transformations of scaling, translation and rotation, are used as test images.

Figures 2.10–2.12 show the transformation detection results for the Lena image. In both these cases, the training images are the non-transformed original and transformed Lena images of dimensions 128×128 and 256×256 . Whenever it comes to scaling detection, a 256×256 times transformed image of Lena is used as the test image, whereby the MLP architecture [52,53] is able to successfully detect the scaling transformation as shown in Fig. 2.10 as a scaling of 2.0079. In Fig. 2.11,

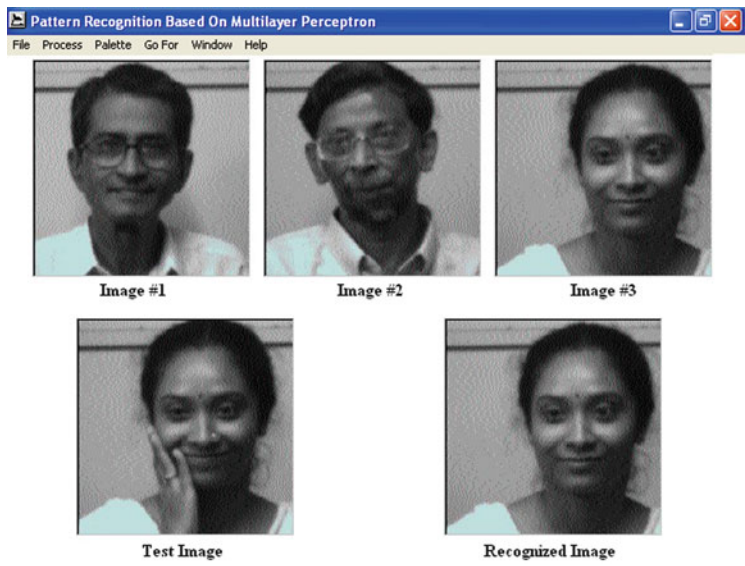


Fig. 2.9 Real-life image recognition

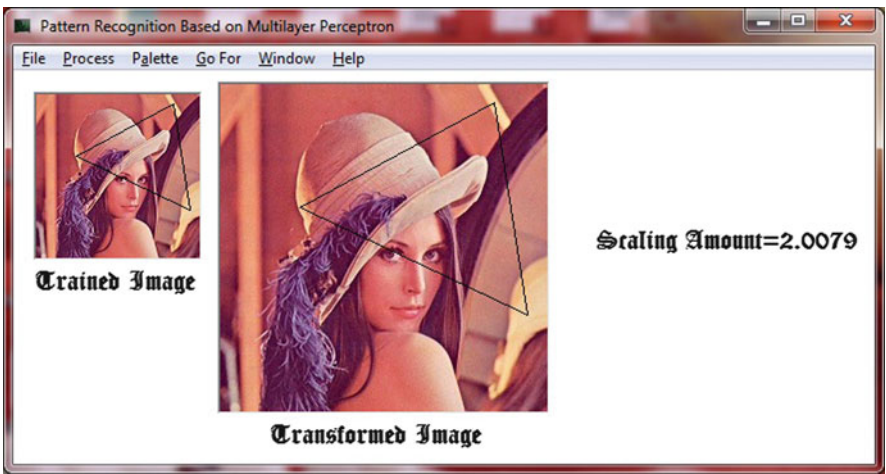


Fig. 2.10 Scaling detection results for Lena image

the same 128×128 size Lena image is used as the test image but with an orientation of 90° clockwise. The MLP architecture [52, 53] successfully detects the rotation transformation as an angle of 89.963° . Figure 2.12 shows the detection capabilities of the MLP architecture [52, 53] when it comes to the detection of a combination of rotation and scaling of the Lena image. Here a 256×256 90° clockwise rotated transformed image is used as the test image which when compared to the trained

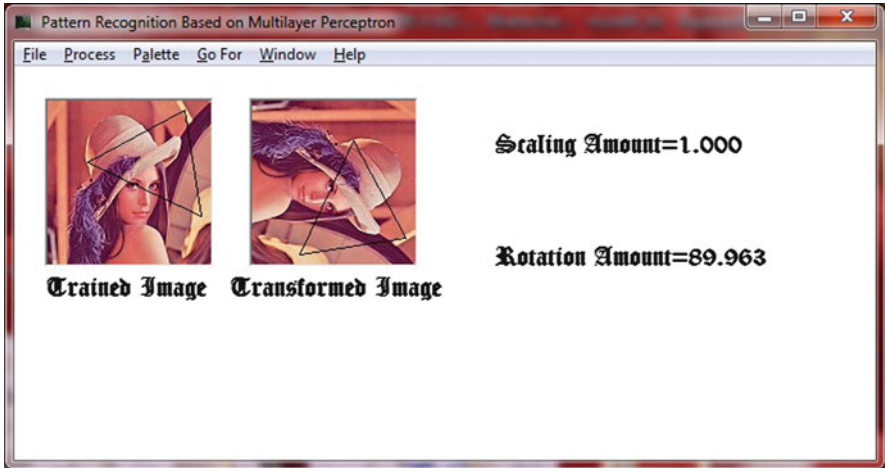


Fig. 2.11 Rotation detection results for Lena image

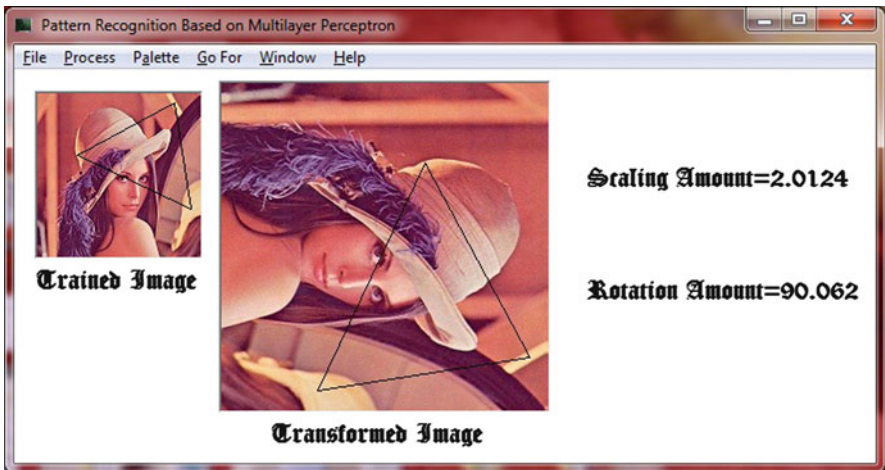


Fig. 2.12 Scaling and rotation detection results for Lena image

original non-transformed Lena image, yields a scaling amount of 2.0124 and a rotation amount of 90.062° by means of the trained MLP architecture [52, 53].

Figures 2.13–2.16 show all the possible transformation detection results for the Plus image. Similar to the Lena image, the training image dataset comprises the non-transformed original as well as the transformed Plus images of different dimensions. During scaling detection, a 1.5 times transformed image of Plus is used as the test image, whereby the MLP architecture [52, 53] detects the scaling transformation amount of 1.5162 as shown in Fig. 2.13. For translation detection, a Plus image translated along the x axis by 23 pixels is used to test the trained MLP

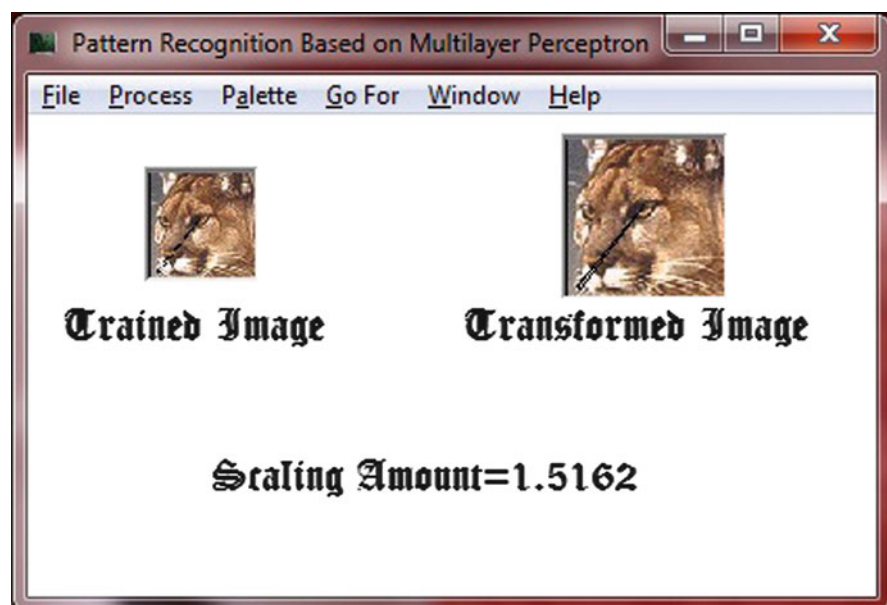


Fig. 2.13 Scaling detection results for Plus image

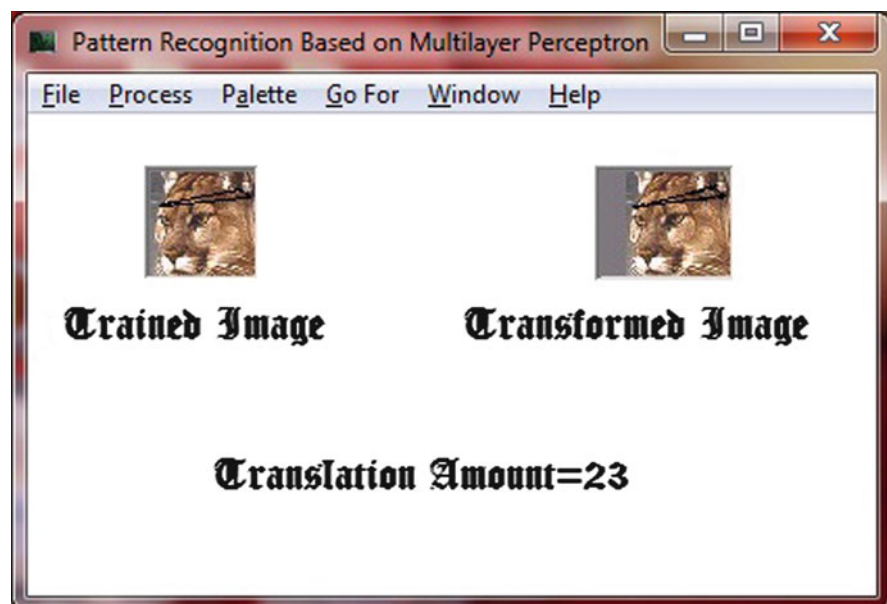


Fig. 2.14 Translation detection results for Plus image

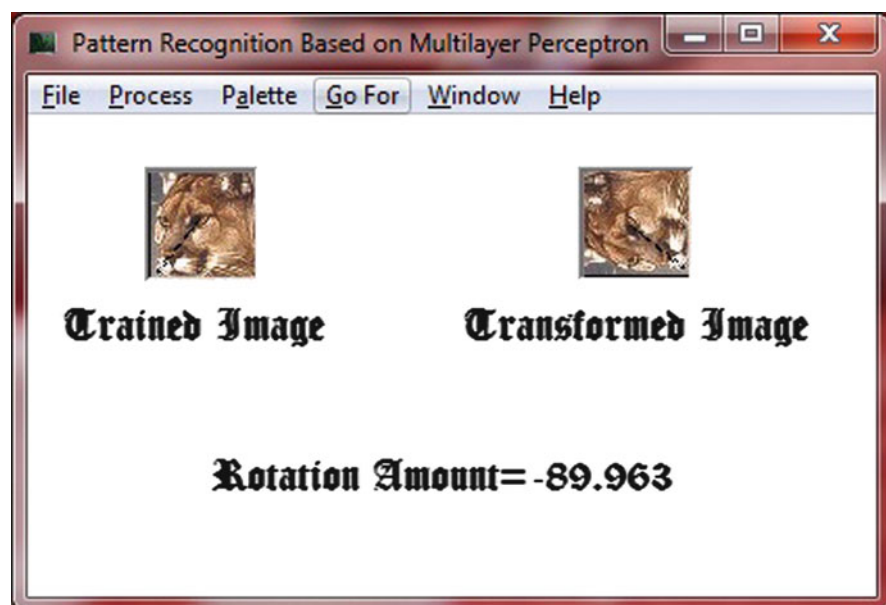


Fig. 2.15 Rotation detection results for Plus image

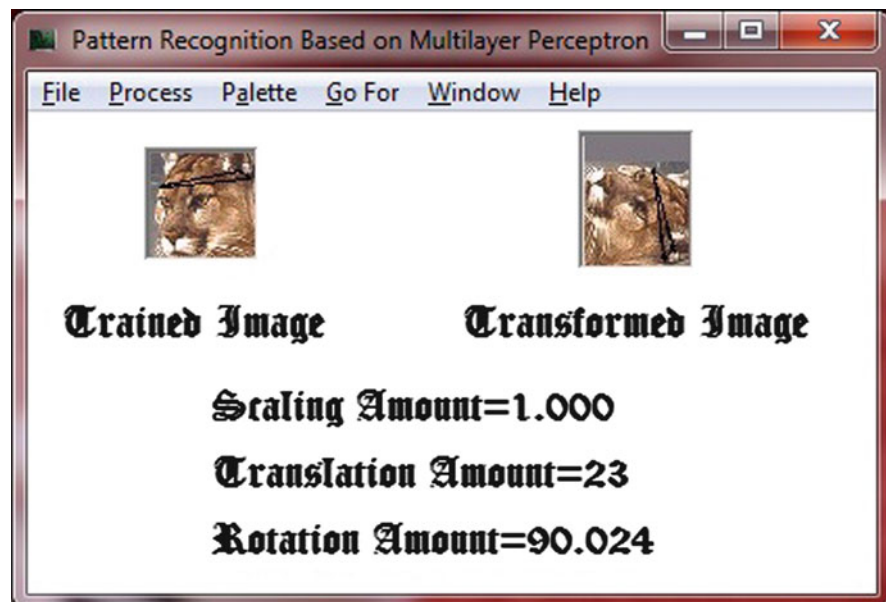


Fig. 2.16 Scaling, translation and rotation detection results for Plus image

architecture [52, 53] as shown in Fig. 2.14, whereby the MLP architecture [52, 53] successfully detects the translated amount. Similarly, the MLP architecture [52, 53] successfully detects the rotation transformation of the Plus image as an angle of -89.963° , i.e., 89.963° anticlockwise. Figure 2.15 shows the rotation detection capability of the MLP architecture [52, 53]. Finally, Fig. 2.16 shows the situation where all the possible transformation detections are made possible by the trained MLP architecture [52, 53]. These transformations include scaling, translation and rotation. For this purpose, a non-scaled (with a scaling of 1.000) 90° clockwise rotated and 23 pixels translated Plus image is used as the test image and presented to the trained MLP architecture [52, 53]. The architecture which is accustomed to all sorts of image transformations by way of training is able to successfully detect the inherent transformations present in the Plus image. The situation is shown in Fig. 2.16, where the different amounts of transformations as measured out by the MLP architecture [52, 53] are also shown alongside. The measured values are (i) scaling of 1.000, (ii) translation of 23 pixels and (iii) rotation of 90.024° .

2.6 Conclusions

This chapter focusses on the image recognition capabilities of a trained multilayer perceptron (MLP) architecture [52, 53]. The generalization and approximation capabilities of an MLP architecture [52, 53] are also illustrated in this chapter with the help of suitable examples. As a sequel, the power of MLP architecture [52, 53] in detecting transformations in test images compared to trained ones is also hinted at using suitable illustrations.

Soft Computing for Image and Multimedia Data
Processing

Bhattacharyya, S.; Maulik, U.

2013, XV, 267 p. 171 illus., 53 illus. in color., Hardcover

ISBN: 978-3-642-40254-8