Assignment 7: Image Classification

1. Create a list with the names called image_files The following list has been made for you to simplify this. Note, in practice you would likely store these files in separate folders and simply read all the files in each folder so that you could add/remove files without editing source code. ['farm1.jpg', 'farm2.jpg', 'farm3.jpg', 'farm4.jpg', 'farm5.jpg', 'farm6.jpg', 'farm7.jpg', 'farm8.jpg', 'city1.jpg', 'city2.jpg', 'city3.jpg', 'city4.jpg', 'city5.jpg', 'city6.jpg', 'city7.jpg', 'city8.jpg', 'desert1.jpg', 'desert2.jpg', 'desert3.jpg', 'desert4.jpg', 'desert5.jpg', 'desert6.jpg', 'desert7.jpg', 'desert8.jpg']

In [17]:
```
import matplotlib.pyplot as plt
import matplotlib.image as mpimg
import numpy as np
from PIL import Image
image_files=['farm1.jpg', 'farm2.jpg', 'farm3.jpg',
             'farm4.jpg', 'farm5.jpg', 'farm6.jpg',
             'farm7.jpg', 'farm8.jpg', 'city1.jpg', 'city2.jpg',
             'city3.jpg', 'city4.jpg', 'city5.jpg', 'city6.jpg',
             'city7.jpg', 'city8.jpg', 'desert1.jpg', 'desert2.jpg',
             'desert3.jpg', 'desert4.jpg', 'desert5.jpg', 'desert6.jpg',
             'desert7.jpg', 'desert8.jpg']
```

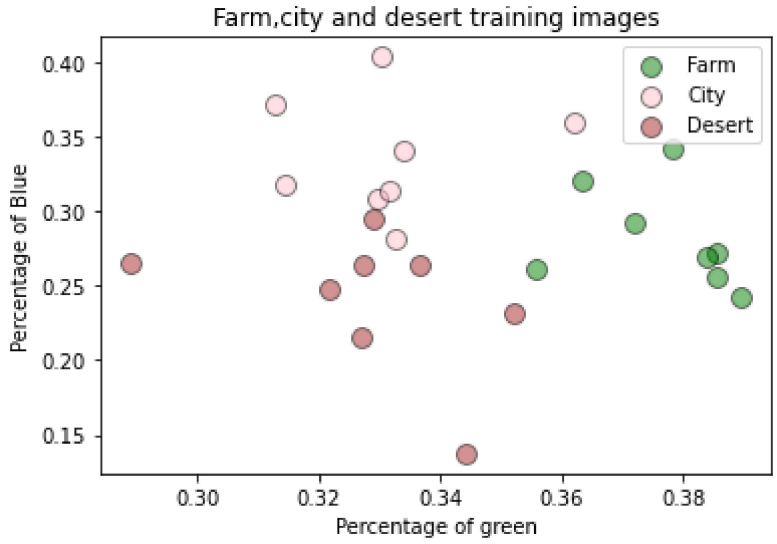
2. Create the scatter plot in the first page Create a scatter plot using the farm/city/desert image data set where each point represents an image, the x axis is the percent green in the image, and the y axis represents the percent blue. The plot should be titled with x and y axis labels, a different color for each of the 3 image classes, and a legend showing which class corresponds to which color. Note: this is with 24 images total, so you will likely want to find some way to automate the process of getting this data for the scatterplot. You can use a for loop to do this. Get the image file name by reading it from the image_files list from step 1. Then perform image processing to get the percent of blue and the percent of green.

In [18]:
```
farm_green=[]
farm_blue=[]
city_green=[]
city_blue=[]
desert_green=[]
desert_blue=[]
```

In [19]:

```
for image in image_files[:8]:
    img=mpimg.imread(image)
    RGBtuple=np.array(img).mean(axis=(0,1))
    averageRed=RGBtuple[0]
    averageGreen=RGBtuple[1]
    averageBlue=RGBtuple[2]
    percentage_green=averageGreen/(averageGreen+averageBlue+averageRed)
    percentage_blue=averageBlue/(averageGreen+averageBlue+averageRed)
    farm_green.append(percentage_green)
    farm_blue.append(percentage_blue)
for image in image_files[8:16]:
    img=mpimg.imread(image)
    RGBtuple=np.array(img).mean(axis=(0,1))
    averageRed=RGBtuple[0]
    averageGreen=RGBtuple[1]
    averageBlue=RGBtuple[2]
    percentage_green=averageGreen/(averageGreen+averageBlue+averageRed)
    percentage_blue=averageBlue/(averageGreen+averageBlue+averageRed)
    city_green.append(percentage_green)
    city_blue.append(percentage_blue)
for image in image_files[16:]:
    img=mpimg.imread(image)
    RGBtuple=np.array(img).mean(axis=(0,1))
    averageRed=RGBtuple[0]
    averageGreen=RGBtuple[1]
    averageBlue=RGBtuple[2]
    percentage_green=averageGreen/(averageGreen+averageBlue+averageRed)
    percentage_blue=averageBlue/(averageGreen+averageBlue+averageRed)
    desert_green.append(percentage_green)
    desert_blue.append(percentage_blue)
```

In [20]:
```
plt.scatter(farm_green,farm_blue,s=100,facecolors='green',edgecolor='black',a
plt.scatter(city_green,city_blue,s=100,facecolors='pink',edgecolor='black',al
plt.scatter(desert_green,desert_blue,s=100,facecolor='brown',edgecolor='black
plt.xlabel('Percentage of green')
plt.ylabel('Percentage of Blue')
plt.title('Farm,city and desert training images')
plt.legend(['Farm','City','Desert'])
plt.show()
```



3. Now create an array of strings called training_target with the category of each. You can use this for convenience: ['farm', 'farm', 'farm', 'farm', 'farm', 'farm', 'farm', 'farm', 'city', 'city', 'city', 'city', 'city', 'city', 'city', 'city', 'desert', 'desert', 'desert', 'desert', 'desert', 'desert', 'desert', 'desert']

In [21]:
```
training_target=['farm','farm','farm','farm',
                 'farm','farm','farm','farm',
                 'city','city','city','city',
                 'city','city','city','city',
                 'desert','desert','desert','desert',
                 'desert','desert','desert','desert']
```

4. Create an empty array of zeros called training_data that will eventually store the percent green and percent blue values. You will be filling this soon. Given the needs of your data set (24 samples and 2 columns), it should have 24 rows and 2 columns.

In [22]:
```
training_data=np.zeros((24,2))
```

5. Now fill the training_data array with the proper values for each image, and observe the values in the array after it is finished. You already have the percent of blue and the percent of green for all images from step 2. Make sure to put those two values in the proper place in the training_data array.

In [23]:
```
for i in range(8):
    training_data[i,0]=farm_green[i]
    training_data[8+i,0]=city_green[i]
    training_data[16+i,0]=desert_green[i]
    training_data[i,1]=farm_blue[i]
    training_data[8+i,1]=city_blue[i]
    training_data[16+i,1]=desert_blue[i]
print(training_data)
```

```
[[0.38537916 0.27250258]
 [0.38947877 0.2416675 ]
 [0.37176749 0.2923693 ]
 [0.38534941 0.25567274]
 [0.38368854 0.26974449]
 [0.37822351 0.34243724]
 [0.35577841 0.26138973]
 [0.36318264 0.32079251]
 [0.33384679 0.33987008]
 [0.31457989 0.31740955]
 [0.32982159 0.30761097]
 [0.33021422 0.40329483]
 [0.31267745 0.37068047]
 [0.3620055  0.35922372]
 [0.33263931 0.28122414]
 [0.33155648 0.31387494]
 [0.28899154 0.26478622]
 [0.32887465 0.29461288]
 [0.32171351 0.24749944]
 [0.35209261 0.23171261]
 [0.32718513 0.21564911]
 [0.33655681 0.2638719 ]
 [0.34419192 0.13749538]
 [0.32732039 0.26438328]]
```

6. Create your classifier. This can often be done in one line. In this case, we suggest using the k-Nearest Neighbors classifier as shown in the tutorial (use k=1), but you can try others if you are interested.

In [24]:
```
from sklearn import neighbors
k1=neighbors.KNeighborsClassifier(n_neighbors=1,weights='distance')
```

7. Train your classifier. Again, this is often only one line of code where you provide the training data and the training target to the classifier you just created to the classifier's 'fit' or 'training' function. For such a small data set, this will be fast, but for larger data sets sometimes this is time consuming. Now you have a trained classifier... great! Now we'll set up the application for it.

In [25]:
```
k1.fit(training_data,training_target)
```

Out[25]:
```
KNeighborsClassifier(n_neighbors=1, weights='distance')
```

8. Now create an empty test_data array and fill it with the proper values for each test image, and observe the filled array and consider if it matches your expectations based on your observations of the images. test_data should start with zeros and be 3 rows (for three test images) and 2 columns (for % green then % blue for each image). Loop through the three test images to fill in the values of the array.

In [26]:
```
test=['test1.jpg','test2.jpg','test3.jpg']
```

In [27]:
```
test_green=[]
test_blue=[]
```

In [28]:
```
for image in test:
    img=mpimg.imread(image)
    RGBtuple=np.array(img).mean(axis=(0,1))
    averageRed=RGBtuple[0]
    averageGreen=RGBtuple[1]
    averageBlue=RGBtuple[2]
    percentage_green=averageGreen/(averageGreen+averageBlue+averageRed)
    percentage_blue=averageBlue/(averageGreen+averageBlue+averageRed)
    test_green.append(percentage_green)
    test_blue.append(percentage_blue)
```

In [29]:
```
test_data=np.zeros((3,2))
```

In [30]:
```
for i in range(3):
    test_data[i,0]=test_green[i]
    test_data[i,1]=test_blue[i]
print(test_data)
```

```
[[0.3269592  0.32688513]
 [0.33429384 0.17936789]
 [0.35004008 0.24578861]]
```

9. Predict the class of the test images. Now predict the classes given the test_data array. This should only take one line of code if the test_data array is prepared.

In [31]:
```
k1_pred=k1.predict(test_data)
print(k1_pred)
```

```
['city' 'desert' 'desert']
```

10. Print the prediction from the test images and compare with the actual images shown below. Make this comparison clear in the output of your code (e.g. prepend with ‘predicted:’ and ‘actual:’). Try to explain any errors if you note any.

In [32]:
```
print("ACTUAL:","city,desert,farm")
print("Predicted: ",k1_pred)
```

```
ACTUAL: city,desert,farm
Predicted:  ['city' 'desert' 'desert']
```

Note: There is a difference between actual and predicted outcome and the predicted outcome. test3.jpg is actually a farm but it has been predicted as a desert