INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

DEVELOPING ADVANCED IMAGE RECOGNITION SYSTEMS WITH CONVOLUTIONAL NEURAL NETWORKS

ANIL KUMAR
Research Scholar
Department of Computer Science and Engineering
J.S.UNIVERSITY, Shikohabad, UP
Dr. K.VIJAYA BHASKAR
Associate Professor
Department of Computer Science and Engineering
J.S.UNIVERSITY, Shikohabad, UP

Abstract

Our everyday lives have been significantly impacted by the ability to browse digital information repositories on the internet and other platforms. A significant amount of this growing body of information consists of digital photos and movies. The task of automatic picture retrieval is made more challenging by the fact that so many photos lack adequate linguistic descriptions. Thus, recovering pictures by analysing their visual information is an intriguing and significant area of science. This issue is resolved by content-based image retrieval (CBIR), which effectively represents a picture's colour, shape, and texture information and retrieves related images using characteristics that have been retrieved. The method used to extract characteristics from the photos has a significant role in deciding how successful a CBIR system is.Face recognition is a challenging problem in many applications because of the large diversity of face photos and the differences in position, occlusion, and lighting conditions. It is harder to have a high recognition rate when there are relatively few training samples available to train the classifier for each topic.Numerous algorithms have been created to enhance the efficiency of face recognition and CBIR issues; nevertheless, because to its high resolution and computational simplicity, local binary pattern (LBP) is becoming more and more popular as a global feature. As a result, it is extensively used in the fields of face recognition, texture classification, and image retrieval for a range of pictures. Over the last several years, a number of LBP variants have been produced that have poor memory, low recognition rates, sensitivity to changes in noise and lighting, and poor resilience to variations in lightning, occlusion, and posture in face pictures. In light of the aforementioned considerations, the purpose of this thesis is to examine how well enhanced LBP performs in image retrieval and face

recognition applications, as well as to investigate potential benefits and overcome challenges associated with using this to develop patterns with high recognition rates and retrieval precision.

1 INTRODUCTION

Multimedia material is produced and consumed in large quantities as a result of the exponential rise of image equipment, such as digital cameras and scanners. All of this information is essentially meaningless without computer-assisted browsing, searching, and retrieval capabilities to find the desired material. Utilising this massive amount of data for a specific purpose requires an efficient system that enables data retrieval based on user interest. For this issue, content-based image retrieval, or CBIR, offers a solution. CBIR is a method that searches a huge picture database according to the user's interests using the visual contents. Through some similarity analysis, it compares the query picture's characteristics with those of the image database [1]. The CBIR system takes one picture as input, and instead of producing a single image, it generates a list of images. Much effort has been done in this area to establish an effective CBIR system [2, 3, 4, 5, 6]. Many researchers have shown a great deal of interest in CBIR in recent years, offering fresh approaches and answers. Low level feature extraction, such as colour, form, and texture, is the foundation of the majority of this field's study [7], [8]. The process's next major phase is feature extraction. Images are used to extract visual and linguistic descriptors, which are then represented in a specific form inside the data space. In order to return the most relevant photos based on certain distance measures, the system lastly computes the distance between the modified characteristics of the query image and all of the images in the database.

Conventional neural networks, or CNN

To perceive things, object recognition systems make use of learning algorithms and extracted characteristics. Many machine learning (ML) and deep learning (DL) methodologies may be used for object identification, e.g. Convolutional neural networks (CNNs), region-based CNNs (R-CNNs), support vector machines (SVMs) [4, 5], etc. Object recognition finds many uses in fields such as biometrics, robotics, defence, visual surveillance, and driving assistance. While surveillance comprises a warning framework for the detection of a pedestrian on pictures collected from infrared cameras [7] and vehicle recognition in the road environment,

ISSN: 2057-5688

driver assistance systems include lane detection frameworks and the identification of impediments before the vehicle. The challenges stem from the several factors affecting the recognition performance, such as variations in lighting, large variations in deformations, shadows, partial occlusion, etc. An attempt is made to work towards the efficient object recognition framework by taking these facts into account. The fundamentals of object identification, deep learning, image decomposition, and thermal imaging are covered in this chapter

Feature extraction is a widely used and significant approach for object recognition. Features are elements or patterns in a picture that help identify an item. Conventional methods in computer vision for feature identification include Scale-Invariant Robust Features (SIFT) [8], Accelerated Robust Features (SURF) [9], and so forth. Conventional feature extraction may be replaced by a CNN [5] as CNNs are powerful enough to extract complex features that reveal the picture in much more detail, become adept at using the explicit features, and are very skilled. At this point, a lot of work has been done.

When it comes to picture feature matching, deep learning-based feature matching uses a deep CNN model that takes the image component into account. While deep learning approaches are better capable of scaling obstacles, rotation, distortion, and other challenges, and have expanded the bounds of what was previously thought possible, this does not imply that standard computer vision processes are no longer relevant.

1.1 Content-Based Image Recovery (CBIR)

Automatic image retrieval has become an urgent research topic due to the handling of terabytes of unlabelled picture data created and digitally stored in large repositories that include visual data both on the web and in the network computing system. Across the globe, these photographs are stored as unstructured multimedia data in several databases, most of which are available online. This growth may be attributed to the ease with which people may use modern consumer electronics gadgets to produce and preserve photos. It is reasonable to assume that this trend will gather more steam in the coming years, placing a great deal of computational pressure on search engine providers such as Google and Bing to make significant investments in their picture search engines. But reliable textural information about these photos is scarce, which complicates the development of effective image retrieval algorithms. Furthermore, since a human annotator can overlook a crucial picture description or disagree with another person on how to interpret a particular image, consistency can be a serious problem in manually annotated collections [1], [2].

To overcome these challenges, CBIR was created in the early 1990s [3], [4], [5]. Google announced Google Goggles, a new content-based search application for Google Android smartphones that allows users to take a picture of a famous site or work of art and search for additional information about it [6, 7]. Machine-interpretable descriptions of an image's physical characteristics, such as colour, texture, form, etc., are generated by CBIR systems. Next, these descriptions may be compared using a similarity metric, also known as extracted characteristics. Next, the CBIR system determines how similar a given query picture is to every image in the archive, and it orders the photos according to how relevant they are to the user's question in decreasing order.

Images are produced at an exponentially increasing rate by a wide range of devices, including scientific experiments, biomedical imaging, home entertainment systems, military reconnaissance and surveillance flights, civilian and defence satellites, mugshot and fingerprinting equipment, and home entertainment systems. The primary activities of CBIR encompass a wide range of application fields, such as medical imaging, managing art galleries and museums [8], designing architecture and engineering projects, law enforcement and criminal investigations [9], managing trademark and copyright databases [10], managing earth resource resources through remote sensing and other methods [11], managing scientific databases, weather forecasting [12], apparel and fabric design [13], retailing [14], medical imaging [15], and picture arc. The recent focus on multimedia systems has drawn attention to CBIR among scholars. The fundamental components of the CBIR system are shown in Figure 1.1 [16].

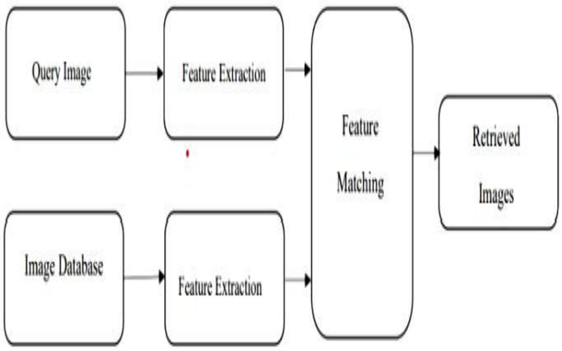


Figure 1 shows the basic CBIR system's flow diagram.

One of the key parts of a CBIR system is the feature extraction block. The feature extraction block uses features to describe a picture, making

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

it possible to find images that are similar to a particular image. A feature captures a particular visual aspect of a picture. Colour, texture, form, and other qualities are the ones that are used in CBIR applications the most. They fall into two broad categories: those that are machine learning-based and those that are handmade. In order to improve the performance of the CBIR system, several recently generated feature descriptors have been made available recently [17], [18], and [19]. In CBIR, a variety of deep learning-based techniques are also used [20], [21], [22]. Images are represented as features using characteristics from colour, texture, and other components. On the picture, these characteristics may be computed both globally and locally.

2 LITERATURE REVIEW

Numerous enhanced patterns have been generated since the revelation of LBP in 1994 [11, 12, 13, 14]. Gathering the intricate features of the pictures is quite difficult, therefore getting strong recall and accuracy for a range of datasets remains a difficult task. Consequently, it's critical to study local patterns and comprehend how they are designed. The purpose of this thesis is to look at a new strategy that improves the traditional LBP's discriminative power for better results. The work that makes up the thesis is summed up as follows:

The advancement of CBIR technology has led to the use of 3D centre symmetric and local mesh ternary patterns for image retrieval.

As LBP and CBIR technologies have advanced, the main challenge now is locating the relevant photos among the many datasets. Due to its challenges in storing the connection between the intensities of nearby pixels and collecting multiscale information in many directions, LBP is still unable to discriminatively capture the local information of the picture. In order to enhance CBIR performance, 3D centre symmetric and bidirectional local mesh patterns (3D-CSLTP, 3D-BLMTP) are being researched. To create multi-scale pictures, they use three separate Gaussian filters with varying standard deviations as smoothing filters. Three R, G, and B planes in a colour picture are regarded as three multiple images, and they are used to create a three-dimensional image. The 3D picture is sliced in five distinct ways to get multidirectional information. It is possible to extract complicated structural arrangements of intensity values in images by generating three

BLMTP in five directions. HOG is included to increase the visibility of 3D-BLMTP and to capture the pixel shape information. It has been combined with another suggested 3D-CSLTP descriptor to improve discriminative power. Retrieving various types of pictures is aided by integrating patterns in different orientations. Using five datasets of face and medical pictures, the method's performance was examined in terms of accuracy, recall, and F-score.

A CBIR method using the colour and texture properties of an image subregion was presented by Vimina et al. (2013). By segmenting the picture into fixed pieces after the use of morphological dilation, this process is therefore carried out in order to identify the Regions of Interest (RoI). According to Nazir et al. (2018), the quantized histogram HSV colour space is used to calculate the colours of the ROIs, and the Grey Level Co-occurrence Matrix (GLCM) is used to compute the texture characteristics. After that, the same amount of Region of Interests are used to compare the Query and the target Images (Memon et al. 2017).

A CBIR model was presented by Ashraf et al. (2018) in relation to colour and Discrete Wavelet Transforms (DWT). The low-level features based on colour, texture, and form has been used to extract comparable pictures (Nazir et al., 2018). It has been noted that these characteristics are important to the extraction processes. Here, the colour characteristics from both the training and query pictures have been retrieved using the colour spaces RGB and YCbCr attributes (Varish et al. 2020). The correct colour contents were extracted by the researchers using the RGB photos that had been converted. Without sacrificing colour accuracy, the converted YCbCr product is used to separate light from colour content. Colour characteristics were adopted and utilised in conjunction with the DWT and histogram to reduce computing operations and improve the efficiency of the search process. Additionally, pictures from the trained image database were retrieved using the Artificial Neural Network (ANN) (Perumal et al. 2017).

Mistry et al. (2017) used the hybrid features and distance measurements to conduct a research on the CBIR. The present study included the integration of hybrid feature types with several feature descriptors, including spatial features, frequency domain features, and Gabor Wavelet Transform. Additionally, the Colour and Edge Directivity Descriptors (CEDD) and Binarized Statistical Image Features (BSIF) were used to boost the retrieval system's efficiency. The HSV colour histogram, the colour moment, the BSIF, the CEDD, and so on were ideally used to recover features. The colour quantization and colour space conversions were discovered to be the characteristics that could be correctly recovered with the use of the HSV histogram. Using the BSIF to extract features involves both converting an RGB picture to grayscale and choosing a patch from the grayscale image. Several tests were run using the presented methodology, and the results showed how effective the suggested approach was when compared to the approaches that were already in use.

In 2016, Zheng et al. presented a hybrid method for picture retrieval. The photos were initially sent to the Hue-Saturation-Intensity (HSI) colour space in order to carry out the feature retrieval phase. After that, the images were divided into blocks. After that, the main block was correctly recovered, and the Robert edge operator was used to create two vertical and horizontal histograms. The matching methods were carried out using the weighted Euclidean distance criteria, and the weights were thereafter empirically determined without any specific consideration of their adjustment levels. When compared to the current methods and the non-adjusting weights, the Low accuracy factor was evidently a weakness of this system.

Through the integration of the Local Base Pattern (LBP) methodology for texture and the Colour Information Feature (CIF) technique for colour feature representation, Liu et al. (2017) proposed a method to distinguish between pictures and search inside an image. Thus, the LBP retrieves the textural feature in order to generate the picture descriptor. Thus, a vast database is used to facilitate a variety of experiment kinds, and the results show the improvised performance level in both the extraction and classification of the photos.

Ahmed et al. (2018) used colour and shape characteristics in feature fusion approaches to conduct research on the CBIR. The items in the picture might be correctly identified because to these properties. It has been discovered that the spatial colour feature in the feature vector improves the picture extraction procedure. The colour characteristic has been retrieved by means of the RGB colour factor. Conversely, the object's corners and edges may be recovered using the gray-level picture, which has been accepted (Tian et al. 2008). Finding the component with the largest variance is made easier by doing dimension reduction on the feature vector. It seems that the compact data

ISSN: 2057-5688

vectors are what Bag of Word (BoW) uses as input to quickly index and retrieve pictures.

3 DATASET AND PERFORMANCE METRICS

The purpose of this part is to offer a detailed explanation of the experimental database and performance metrics that were used in order to evaluate the efficiency of our Facial Expression Recognition (FER) system.

3.1 Bosphorus

This three-dimensional collection (Savran et al. 2008) includes a broad variety of facial expressions and a wide range of situations that people find themselves in. Attempts are often made by actors and actresses to produce a depiction that is more authentic. It consists of

- There are six core emotions, a state of neutrality, and a selected action unit that are used to classify the expressions.
- A wide variety of head postures are included in the collection.
- A number of different methods for hiding one's face are shown.

Description

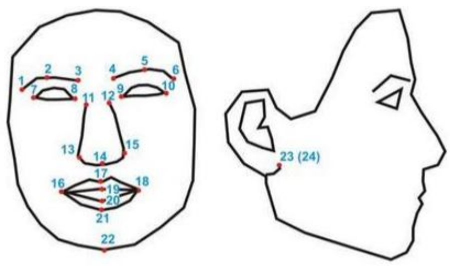
When it comes to the Bosphorus Database, there are 105 distinct problems that are handled utilizing a wide range of idioms. Because of the favorable lighting circumstances, the process of acquiring the 3D face scans was made easier. Some of the most important uses of this technology are the identification of faces and facial expressions. There are features associated with self-occlusion that are included in the face data. A list of the Bosphorus Database's entries is shown in the following.

In contrast, the two-dimensional color photographs, which include high-quality cropped face portions, are stored in a png file, while the three-dimensional data is saved in a bnt file. In the lm2 and lm3 files, the information on twenty-four landmarks in the face is stored in a distinct location.

We manually identified the face landmarks on the two-dimensional photographs to determine whether or not the facial scan produced clear clarity. On the other hand, we used 3D-to-2D correspondences to figure out the coordinates of the three-dimensional landmarks. A three-dimensional scan of the face was used to acquire the annotated landmarks that are shown in Figure 3.1. Names of the landmarks are as follows, in alphabetical order: The table 3.1.

INTERNATIONAL JOURNAL OF
TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688



(Source: http://bosphorus.ee.boun.edu.tr/Home.aspx)

Figure3.1

24 Labelled Landmarks

Table 3.1

Labels of 24 Landmarks

1. Outer left eye brow	2. Middle of the left eye brow
3. Inner left eye brow	4. Inner right eye brow
5. Middle of the right eye brow	6. Outer right eye brow
7. Outer left eye corner	8. Inner left eye corner
9. Inner right eye corner	10. Outer right eye corner
11. Nose saddle left	12. Nose saddle right
13. Left nose peak	14. Nose tip
15. Right nose peak	16. Left mouth corner
17. Upper lip outer middle	18. Right mouth corner
19. Upper lip inner middle	20. Lower lip inner middle
21. Lower lip outer middle	22. Chin middle
23. Left ear lobe	24. Right ear lobe

The BNT file contains a substantial quantity of information over its whole. The data size, the number of rows and columns, the minimum depth value, the length of the name of the two-dimensional image file, and the format of the file (png) are some of the features that are contained in this section.

Constructing object files may be accomplished via the use of Bnt files, which are made up of the v, vt, and f components. When it comes to coordinates, the letter v stands for a three-dimensional coordinate, the letter vt represents a two-dimensional value that has been normalized, and the letter f indicates facet values.

4. Characteristics of the Local Binary Pattern

Determined by Depth

When converting a picture into an array or image of integer labels that describe the image's look on a smaller scale, the LBP operator is used. This operator is used to convert the picture. For the purpose of carrying out further picture analysis, these labels or their statistics, which are commonly represented by the histogram, are used. According to the findings of Sheng et al. (2016), LBP is a feature that is widely used in conditions involving categorization. The portions of the face that undergo the most significant changes as a result of facial expression are the areas that are located around the eyes and the mandible. In light of this, the landmarks, which consist of the outer left eye brow, the middle left eyebrow, the inner left eye brow, the inner right eye brow, the middle right eye brow, the outward right eye brow, the left mouth corner, the upper lip outer middle, the lower lip outer middle, and the right mouth corner, are taken into consideration respectively. The subsequent step involves the use of a 5x5 window that is based on LBP-based depth values in order to extract features that are located in close proximity to the 10 unstable feature locations. The end result is the generation of a feature set consisting of 250 features, with 25 features being assigned to each major landmark. To show the method for depth features, which is based on LBP, the following image is provided.

We are operating under the assumption that the norms have a value of 0.1 for the ℭ parameter, and the vector v is the non-normalized vector that includes all of the histograms that are included inside a particular block. Typically, the size of photographs that fall within this range are 128 pixels by 64 pixels. In order to determine the units' dimensions, a grid that is 8 by 8 dimensions is used. The traditional measurement for a cell is two by two, and this is the measurement that is used. In the event that the gradient signals are disregarded, there are nine divisions that are equally distributed from zero to one hundred and eighty degrees. In terms of size, the HOG descriptor is pretty substantial, taking about 18720 bytes.When contrasted with the HOG features that were obtained from the grayscale picture, the HOG features that were created from the range image were seen to be substantially more effective.

4.1 The characteristics of local binary patterns that are derived from depth

With the help of the LBP operator, you may extract an image and get either an array or a picture of integer labels that describe the appearance of the image on a smaller scale. Furthermore, these labels or their statistics, most often the histogram, are used for the purpose of additional image analysis. One of the characteristics that is often used in the process of classification is low back pain (LBP) (Sheng et al. 2016).

The eyes and the jaw are the areas that contribute the most to the deformation that happens while forming an emotion. It is for this reason that the landmarks are taken into consideration. These landmarks include the following: the outer left eye brow, the middle left eyebrow, the inner left eye brow, the inner right eye brow, the middle right eye brow, the outer right eye brow, the left mouth corner, the upper lip outer middle, the lower lip outer middle, and the right mouth corner. In the next step, a 5x5 window of LBP-based depth values is used in order to extract features that are located around the 10 unstable feature locations. Twenty-five features are given to each landmark, bringing the total number of features created to 250. An example of the algorithm for LBP-based depth features is shown in the following graphic.

Algorithm:

Input: Depth values

Output: LBP based depth features

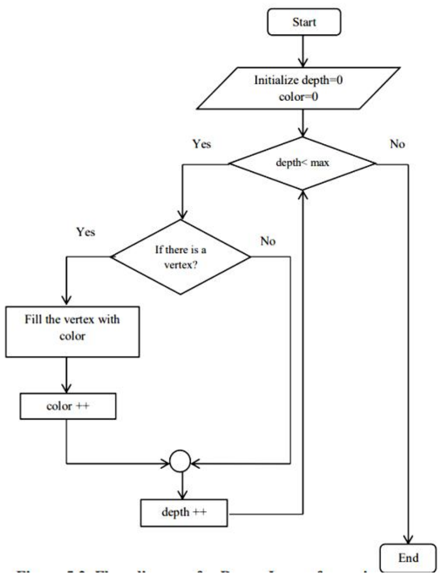
1. Locate the unstable land marks.
2. A5*5 window of depth values is extra points.
3. For each pixel, find its 8 neighbors and th values of pixels in the 3*3 window with tl
4. If the average value is less than the neigl 1. This yields us an 8-bit binary number.

5 Deep Learning For 3D Face Expression Recognition In Highlighted Images

When it comes to any aspect of your life, you will never be able to avoid the influence of contemporary technology. In order for this shift to take place, the two technologies that are most crucial are deep learning and machine learning. In the same way as the industrial revolution came before them, these forces that initiate change have an impact on the places where people live and the ways in which they communicate. These relatively new technologies are already having an influence on our day-to-day lives in a variety of ways, including the improvement of health care, the construction of electrical infrastructure, the increase of agricultural yields, the enhancement of gadget intelligence, and the monitoring of climate change.

Methodology for the Creation of Range Images

Figure4.1shows the flow diagram that represents how to construct range images.



ISSN: 2057-5688

Figure 5.1 Formation of Range Images: A Flow Diagram

ISSN: 2057-5688

5.4

CREATING FEATURED IMAGES

The raw face photos were not utilized in the process of learning the facial characteristics; rather, we used two highlighted images since these images reflect some of the most prominent traits at the beginning of the process. An explanation of the process of the production of two featured images may be found in the following parts.

Evaluation of the size of the edge's direction

It is possible to calculate the edge direction magnitude for each pixel in the range picture by using the Kirsch masks, K z, in each of the eight distinct compass directions, as shown in Figure 5. 2

ij

NW		N		NE
	-3 5 5 -3 0 5 -3 -3 -3	-3 -3 5 -3 0 5 -3 -3 5	-3 -3 -3 -3 0 5 -3 5 5	
W	5 5 5 -3 0 -3 -3 -3 -3	Pixel	-3 -3 -3 -3 0 -3 5 5 5	E
	5 5 -3 5 0 -3 -3 -3 -3	5 -3 -3 5 0 -3 5 -3 -3	-3 -3 -3 5 0 -3 5 5 -3	
SW		S		SE

Figure 5.2 The eight compass directions are used by Kirsch masks.

Figure 5.4 illustrates the results obtained by applying the Kirsch edge mask in each of the eight directions of the compass. Additionally, the amplitude of an edge is shown in each direction.

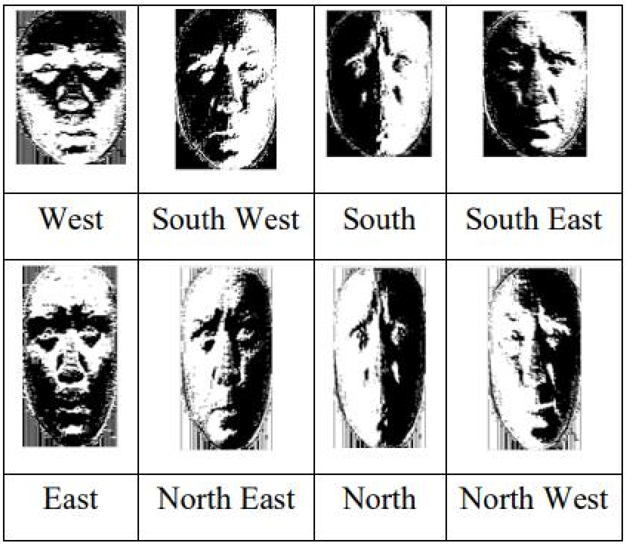


Figure5.4 The eight compass orientations yielded Kirsch edge replies

Identifying the magnitude of the highest edge direction

Equation (5.1) is used to calculate the maximum edge direction magnitude, E, using the provided picture I, which has a size of m * n.

$$E_{i,j} = \max_{z=1,2,\ldots,8} \sum_{l=-1}^{1}\sum_{k=-1}^{1} K_j^z * I_{l+i,k+j} \qquad (5.1)$$

the range of values for i is 2–m—1 and for j is 2–n—1. Eight Kirsch kernels, one for each of the four cardinal directions, provide the basis of the calculation. That is mask-dependent; it must be the one that generates the largest edge magnitude.

5.4.1 LDP image computation

Kirschi is a notation that may be used to denote the outcomes of applying any one of the eight Kirsch Kernels to any integer i between 1 and 8. Assume that the greatest edge magnitude that might be accomplished is denoted by the symbol Kirschmax. In order to determine the LDP value for each pixel, the equation (5.2) is used.

$$LDP = \sum_{i=0}^{7} b(Kirsch_i - Kirsch_{max}) * 2^i$$

$$\text{where} \quad b(x) = \begin{cases} 1; & x > 0 \\ 0; & otherwise \end{cases} \qquad (5.2)$$

There is a sturdy featured picture produced as a consequence of the image being developed by taking into consideration all eight directions.

The Creation of Local Binary Patterns (LBPs) in Images

As we discussed earlier in Section 4.1.3, the LBP image operator has the capability to turn a picture into either a label array or an image that contains numbers. In order to do further analysis of the image, you may make use of these labels or related statistics, such as the histogram. Because of this circumstance, the LBP approach has to be modified.

In order to segment the input image, several blocks of a fixed size are used instead. When we look at a window that is three by three, we are able to view the eight pixels that are on each side of each pixel, including the top, bottom, left, and right pixels. As shown in Figure 5.6, rather than comparing the value of the center pixel to the intensity values of the eight neighbors, as shown in Figure 5.5, we compute the average intensity values of the neighbors and then compare those values to the intensity values of the eight neighbors. This is illustrated in Figure 5.6. It is deemed to be one after the conditions have been satisfied; otherwise, it is thought to be zero. A feature is a binary number of eight bits, which is the outcome of the operation. Regarding the processing of the neighbors, it

ISSN: 2057-5688

is possible to proceed in either a clockwise or anticlockwise way.

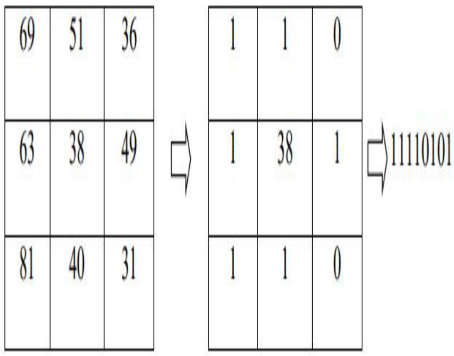


Figure 5.5NormalLocalBinary Pattern

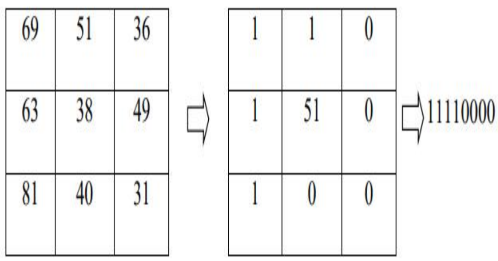


Figure5.6 Local Binary Pattern Modification

A replacement for the features of the pictures, which are mostly local primitives, is provided by the LBP codes, which are binary integers that have been generated from binary numbers. They include a broad range of rounded edges, patches, flat parts, and other characteristics that are comparable to one another. In the past, instructions were given on the network.Deep Neural Networks (DNNs) have been successful for analytical tasks as a result of their ability to transform sparse low-level input into rich high-level features. This ability is the reason for their rapid rise in popularity. Deep neural networks have a lot of drawbacks, one of which is the vast number of network parameters that need to be calculated. This is despite the fact that deep neural networks are superior than shallow neural networks when it comes to the process of converting low-level information into high-level features. On the other hand, this limitation may be circumvented by using deep CNN models that have already been pre-trained and are available for commercial acquisition. These deep CNN models, which are already trained and are accessible for commercial usage, are able to solve recognizing issues in a way that is both consistent and successful. This is because of the generalizability of these models. Utilizing a transfer learning methodology that makes use of a pre-trained model that has been fine-tuned is the most effective method to take when dealing with recognizing issues. This is considered to be the greatest way to adopt.

CONCLUSION

An overview of object recognition is given in this chapter along with a block diagram and an explanation of each component. There is additional discussion of object recognition applications in the many disciplines. Furthermore, the difficulties in the area of object identification have also been thoroughly examined. Furthermore, a thorough explanation of DL models and the fundamentals of deep learning has also been completed. Additionally, the ideas of thermal imaging systems and picture decomposition have been discussed. Next, it was addressed what motivated the researcher to do this study. Finally, at the conclusion of the chapter, the goals of the study and the structure of the thesis are outlined.

REFERENCES

[1] In Wang X.-Y. & Chen Z.-F., "A rapid fractal coding in image retrieval application," Fractals, vol. 2009, 17, no. 04, pp. 441–450.

[2] Liu Y., Zhang D., Lu G., Ma W.-Y., "A survey of high-level semantics content-based image retrieval," Pattern Recognition, vol. 40, 2008, pp. 262-282.

[3] In the International Journal of Medical Informatics, volume, Müller H., Michoux N., Bandon D., and Geissbuhler A. reviewed content-based image retrieval systems in medical applications, discussing their therapeutic advantages and future prospects. No. 73, pages 1–23, 2004.

[4] "A survey of dictionary learning algorithms for face recognition," Xu Y., Li Z., Yang J., and Zhang D., IEEE access, vol. 5, 2017, pp. 8502–8514.

[5] Meyer-Baese A. & Schmid, V. J., Signal analysis and pattern identification in medical imaging. 2014 Elsevier.

[6] "Towards large-scale histopathological image analysis: Hashing-based image retrieval," Zhang X., Liu W., Dundar M., Badve S., and Zhang S., IEEE Transactions on Medical Imaging, vol. 34, no. 2, 2014, pp. 496-506.

[7] "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," Ojala T., Pietikäinen M., and Mäenpää T., IEEE Transactions on

pp. 971–987 in Pattern Analysis & Machine Intelligence, no. 7, 2002.

[8] Dalal N. "Histograms of oriented gradients for human detection," by Triggs B. and others, in IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1. IEEE, pp. 886–893, 2005.

[9] "Imagenet classification with deep convolutional neural networks," by Krizhevsky A., Sutskever I., and Hinton G. E., Advances in Neural Information

Processing Systems, vol. 25, pp. 1105, 1197 (2012).

[10] Simonyan K. & Zisserman A. (2014) arXiv preprint arXiv:1409.1556, "Very deep convolutional networks for largescale image recognition."

[11] Srivastava P. and Khare A., "Content-based image retrieval using multiscale local binary pattern," Multimedia Tools and Applications, vol. 77, no. 10, pp.12-377-803, 2018.

[12] Verma M. & Raman B. "Local binary co-occurrence pattern with centre symmetry" for biomedical image retrieval, texture, and face recognition," Journal of Visual Communication and Image Representation, vol. 32, 2015, pp. 226-236.

[13] Tan X and Triggs B., IEEE Transactions on Image Processing, vol., "Enhanced local texture feature sets for face recognition under challenging lighting conditions." 19,no. 6, 2010, pp. 1635–1650.

[14] Murala S. and Wu Q. J., "Retrieval of MRI and CT images using local ternary co-occurrence patterns: A novel feature descriptor," Neurocomputing, vol. 119, pages 399–412, 2013.

[15] Singhal A., Lall B., and Agarwal M. "Texture, face, natural, and biomedical image retrieval using 3D local ternary co-occurrence patterns," Neuro computing, vol.313, pages 325–345, 2018.

[16] "3D local circular difference patterns for biomedical image retrieval," Mohite N., Waghmare L., Gonde A., and Vipparthi S., International Journal of Multimedia Retrieval of Information, vol. 8, no. 2, 2019, pages 115–125.

[17] According to Sucharitha G. & Senapati R. K., "Local directional edge binary patterns and zernike moments for biomedical image retrieval," Multimedia Tools and Applications, vol. 2020; 79, no. 3, pp. 1847–1864.

[18] Turk M. "Eigenfaces for recognition," by Pentland A. and others, Journal of Cognitive Neuroscience, vol. 3, no. 1, 1991, pp. 71–86.

[19] Belhumeur P. N., Hespanha J. P., Kriegman D. J., IEEE Transactions on Pattern Recognition, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection."Machine Intelligence and Analysis, vol. 19, no. 7, 1997, pp. 711–720.

[20] Belkin, M. , Niyogi P., "Laplacian eigenmaps for data representation and dimensionality reduction," Neural Computation, vol. 15, no. 6, 2003, pp. 1373–1396).

[21] Shivashankar S. "PCA plus LDA on wavelet cooccurrence histogram features: Application to CBIR," by & Veerashetty S. K. S., in International Workshop on Multi-disciplinary Trends in Artificial Intelligence. 2011; Springer, pp. 193-200.

[22] Ma Liao. and Gu X., "Discriminative feature learning and subspace extended sparse representation as a face recognition approach," Neuro computing, vol. 373, 2020, pp. 35–49.

[23] N.-S. Vu, "Investigating gradient orientation and magnitude patterns for facial recognition," IEEE Transactions on

Information Forensics and Security, vol. 8,no. 2, 2012, pp. 295–304.
[24] Nguyen H.-T. Local gradient patterns for facial identification, by & Caplier A., IEEE Transactions on Information Forensics and Security, vol. 10, no. 8, pages 1739–17401751, the year 2015.
[25] Karanwal, S. and Diwakar M., "LBP for face recognition based on neighbourhood and centre differences," Pattern Analysis and Applications, pp. 1–21, 2021.
[26] In KSII Transactions on Internet, Yang J., Zhang L., Wang Y., Zhao T., Sun W., and Park D. S., "Face recognition based on weber symmetrical local graph structure."as well as Information Systems, vol. 12, no. 4, 2018, pp. 1748–1759.
[27] http://www.oasis-brains.org/, the OASIS-MRI database, [Online; accessed 21-FEB2023].