Survey of Content Based Image Retrieval using Local Binary Patterns

Hemantkumar B. Jadhav, Asst. Professor, VACOE, Ahmednagar, India, Hemantvacoe@gmail.com

Amit Jadhav, PG Student , VACOE, Ahmednagar, India, amitjadhav184@gmail.com

Abstract: This paper, describes a survey of using local binary patterns for feature extraction. It focuses on a new coding scheme, local Gabor maximum edge position octal patterns (LGMEPOP) which is proposed for content based image retrieval. The standard local binary pattern (LBP) collects the sign edge (binary code) information between the center pixel and its surrounding neighbors in an image. Further, the concept of LBP is extended to local maximum edge binary pattern (LMEBP) which collects the sign code (binary code) using the magnitude edges. These magnitude edges are collected based on the maximum edges between the center pixel and its surrounding neighbors in an image. The octal code is coded based on the maximum edge positions (MEP) on Gabor responses. Specially, each pixel of every Gabor response gains eight edges based on the relationship between the referenced pixel and its neighbors. LGMEPOP utilizes the first three dominant (maximum) edge positions in an octal code generation. Then, these three maximum edge positions are encoded into three-eight octal numbers to produce the LGMEPOP. Further, the LGMEPOP is classified into two categories which are named as sign maximum edge position octal pattern (SMEPOP) and magnitude maximum edge position octal pattern (MMEPOP)

Keywords —CBIR, Text, Color, Shape, Local Binary Pattern Introduction

I. INTRODUCTION

Content-based image retrieval, a technique which uses visual contents to search images from large scale image databases according to users interests, has been an active and fast advancing research area since the 1990s. During the past decade, remarkable progress has been made in both theoretical research and system development [1]. However, there remain many challenging research problems that continue to attract researchers from multiple disciplines. Researchers from the communities of computer vision, database management, human-computer interface, and information retrieval were attracted to this field. Since then, research on content-based image retrieval has developed rapidly [2-7].

Early techniques were not generally based on visual features but on the textual annotation of images. In other words, images were first annotated with text and then searched using a text-based approach from traditional database management systems. Text-based image retrieval uses traditional database techniques to manage images. Through text descriptions, images can be organized by topical or semantic hierarchies to facilitate easy navigation and browsing based on standard boolean queries. However, since automatically generating descriptive texts for a wide spectrum of images is not feasible, most text-based image retrieval systems require manual annotation of images. Obviously, annotating images manually is a cumbersome and expensive task for large image databases, and is often subjective, context-sensitive and incomplete. As a result, it is difficult for the traditional text-based methods to support a variety of task-dependent queries.

As a result of advances in the Internet and new digital image sensor technologies, the volume of digital images produced by scientific, educational, medical, industrial, and other applications available to users increased dramatically. The difficulties faced by text-based retrieval became more and more severe. The efficient management of the rapidly expanding visual information became an urgent problem. This need formed the driving force behind the emergence of content-based image retrieval techniques.

Content based image retrieval uses the visual contents of an image such as color, shape, texture, and spatial layout to represent and index the image. Among very popular local image descriptors which has shown interesting results in extracting soft facial biometric traits is the local binary patterns (LBP) [8,12,19,20]. LBP can be seen as statistics of labels computed in the local pixel neighborhoods. The LBP method describes each pixel's neighborhood by a binary code which is obtained by first convolving the image with a predefined set of linear filters and then binarizing the filter responses. The bits in the code string correspond to binarized responses of different filters. The LBP-like methods showed very good results in different computer vision tasks, including nontraditional texture problems such as face recognition, gender classification, age estimation and motion analysis [12, 13, and 20]. The LBP method can be seen as a unifying approach to the traditionally divergent statistical and structural models of texture analysis.

Perhaps the most important property of the LBP operator in real world applications is its invariance against monotonic gray level changes caused, for example, by illumination variations. Another equally important property is its computational simplicity, which makes it possible to analyze images in challenging real-time settings. Furthermore, LBP is also very flexible: it can be easily adapted to different types of problems and also used together with other image descriptors. The LBP is related to many well-known texture analysis operators as discussed in [16]. The arrows represent the relations between different methods, and the texts beside the arrows summarize the main differences between them. As shown in [9], local binary patterns can also be seen as a combination of local derivative filter operators whose outputs are quantized by thresholding.

Face analysis is perhaps the most fascinate

ng application of LBP [12]. While texture features have been successfully used in various computer vision applications, only relatively few works have considered them in facial image analysis before the introduction of the LBP based face representations in 2004 [29]. Since then, the methodology has inspired plenty of new methods in face analysis. For instance, it has been successfully applied to face detection [11], face recognition [8], facial expression recognition [21], gender and age classification [24] and head pose estimation [15].

II. Image Content Descriptors

Generally, image content may include both visual and semantic content. Visual content can be very general or domain specific. General visual content include color, texture, shape, spatial relationship, etc. Domain specific visual content, like human faces, is application dependent and may involve domain knowledge. Semantic content is obtained either by textual annotation or by complex inference procedures based on visual content.

A good visual content descriptor should be invariant to the accidental variance introduced by the imaging process (e.g., the variation of the illuminant of the scene). However, the T. Mäenpää, M. Pietikäinen, Texture analysis with local binary patterns, in: C. Chen, P. Wang (Eds.), Handbook of Pattern Recognition and Computer Vision, World Scientific, 2005, pp. 197–216.e is a tradeoff between the invariance and the discriminative power of visual features, since a very wide class of invariance loses the ability to discriminate between essential differences. Invariant description has been largely investigated in computer vision (like object recognition), but is relatively new in image retrieval [11].

A visual content descriptor can be either global or local. A global descriptor uT. Mäenpää, M. Pietikäinen, Texture analysis with local binary patterns, in: C. Chen, P. Wang (Eds.), Handbook of Pattern Recognition and Computer Vision, World Scientific, 2005, pp. 197–216.ses the visual features of the whole image, whereas a local descriptor uses the visual features of regions or objects to describe the image content. To obtain the local visual descriptors, an image is often divided into parts first. The simplest way of dividing an image is to use a partition, which cuts the image into tiles of equal size and shape. A simple partition does not generate perceptually meaningful regions but is a way of representing the global features of the image at a finer resolution. A better method is to divide the image into homogenous regions according to some criterion using region segmentation algorithms that have been extensively investigated in computer vision. A more complex way of dividing an image, is to undertake a complete object segmentation to obtain semantically meaningful objects (like ball, car, horse). Currently, automatic object segmentation for broad domains of general images is unlikely to succeed.

➢ **Color:**

Color is the most extensively used visual content for image retrieval. Its three-dimensional values make its discrimination potentiality superior to the single dimensional gray values of images. Before selecting an appropriate color description, color space must be determined first.

➢ **Texture**

Texture is another important property of images. Various texture representations have been investigated in pattern recognition and computer vision. Basically, texture representation methods can be classified into two categories: structural and statistical. Structural methods, including morphological operator and adjacency graph, describe texture by identifying structural primitives and their placement rules. They tend to be most effective when applied to textures that are very regular. Statistical methods, including Fourier power spectra, co-occurrence matrices, shift-invariant principal component analysis (SPCA), Tamura feature, Wold decomposition, Markov random field, fractal model, and multi-resolution filtering techniques such as Gabor and wavelet transform, characterize texture by the statistical distribution of the image intensity.

➢ **Shape**

Shape features of objects or regions have been used in many content-based image retrieval systems. Compared with color and texture features, shape features are usually described after images have been segmented into regions or objects. Since robust and accurate image segmentation is difficult to achieve, the use of shape features for image retrieval has been limited to special applications where objects or regions are readily available. The state-of-art methods for shape description can be categorized into either boundary-based or region-based methods. A good shape representation feature for an object should be invariant to translation, rotation and scaling.

III. Basic Local Binary Patterns Operator

The LBP operator, introduced by Ojala et al. [19], is defined as a gray scale invariant texture measure, derived from a general definition of texture in a local neighborhood. It is a powerful means of texture description and among its properties in real-world applications are its discriminative power, computational simplicity and tolerance against monotonic gray-scale changes. The original LBP operator forms labels for the image pixels by thresholding the 3×3 neighborhood of each pixel with the center value and considering the result as a binary number. Figure 1 shows an example of an LBP calculation.

The histogram of these $2^8 = 256$ different labels can then be used as a textureT. Mäenpää, M. Pietikäinen, Texture analysis with local binary patterns, in: C. Chen, P. Wang (Eds.), Handbook of Pattern Recognition and Computer Vision,World Scientific, 2005, pp. 197–216. descriptor for further analysis (e.g. classification).

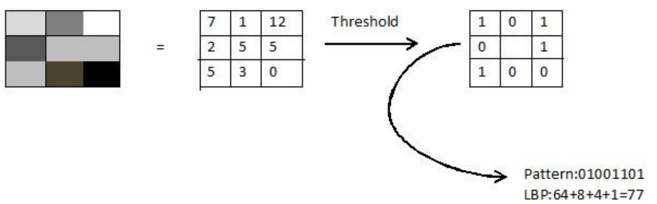

Figure 1: The basic LBP operator.

The LBP operator has been extended to use neighborhoods of different sizes. Using a circular neighborhood and bilinearly interpolating values at non-integer pixel coordinates allow any radius and number of pixels in the neighborhood. The notation (P, R) is generally used for pixel neighborhoods to refer to P sampling points on a circle of radius R. The calculation of the LBP codes can be easily done in a single scan through the image. The value of the LBP code of a pixel (xc , yc) is given by:

$$LBP_{P,R} = \sum_{p=0}^{p-1} s(gp - gc)2^p$$

where gc corresponds to the gray value of the center pixel (xc , yc), gp refers to gray values of P equally spaced pixels on a circle of radius R, and s defines a thresholding function as follows:

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Another extension to the original operator is the definition of so called uniform patterns. This extension was inspired by the fact that some binary patterns occur more commonly than others. A local binary pattern is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or viceversa when the bit patternis traversed circularly. In the computation of the LBP labels, uniform patterns are used so that there is a separate label for each uniform pattern and all the non-uniform patterns are labeled with a single label. This yields to the following notation for the LBP operator:$LBP^{u2}_{P,R}$.

The subscript represents using the operator in a (P, R) neighborhood. Superscript u2 stands for using only uniform patterns and labeling all remaining patterns with a single label. Each LBP label (or code) can be regarded as a microtexton. Local primitives which are codified by these labels include different types of curved edges, spots, flat areas etc. The occurrences of the LBP codes in the image can be collected into a histogram. The classification can then be performed by computing histogram similarities. For an efficient representation, facial images are first divided into several local regions from which LBP histograms are extracted and then concatenated into an enhanced feature histogram for classification.

IV. VARIANTS OF LOCAL BINARY PATTERNS

The success of LBP methods in various computer vision problems and applications has inspired much new research on different variants. Due to its flexibility, the LBP method can be easily modified to make it suitable for the needs of different types of problems. The basic LBP operator has also some problems that need to be addressed. Therefore, several extensions and modifications of LBP have been proposed in the literature with an aim to increase its robustness and discriminative power.

A limitation of the original LBP operator is its small spatial support area. To cope with this problem (i.e. small spatial support area), [16] proposed the multiscale block LBP variant (MB-LBP) which has gained popularity especially in facial image analysis. The key idea behind MB-LBP is to compare average pixel values within small blocks instead of comparing pixel values. The operator always considers 8 neighbors, producing labels from 0 to 255. For instance, if the block size is 3 × 3 pixels, the corresponding MB-LBP operator compares the average gray value of the center block to the average values of the eight neighboring blocks of the same size, thus the effective area of the operator is 9 × 9 pixels. Instead of the fixed uniform pattern mapping, MB-LBP has been proposed to be used with a mapping that is dynamically learned from a training data. In this mapping, the N most often occurring MB-LBP patterns receive labels 0, ... N − 1, and all the remaining patterns share a single label. The number of labels, and consequently the length of the MB-LBP histogram is a free parameter that the user can set.

Wolf et al. [23] considered different ways of using bit strings to encode the similarities between patches of pixels, which could capture complementary information to pixel-based descriptors. The authors proposed a three-patch LBP (TPLBP) and four-patch-LBP (FPLBP) variants. For each pixel in TPLBP, a w × w patch centered at the pixel and S additional patches distributed uniformly in a ring of radius r around it are considered. Then, the values for pairs of patches located on the circle at a specified distance apart are compared with those of the central patch. The value of a single bit is set according to which of the two patches is more similar to the central patch. The code produced will haveS bits per pixel. In FPLBP, two rings centered on the pixel were used whereas only one ring is considered in TPLBP. TPLBP and FPLBP are designed to encode additional texture information compared to LBP.

Tan and Triggs [22] proposed a three-level operator called local ternary patterns (LTP), (using one threshold T) e.g. to deal with problems on near constant image areas. In ternary encoding the difference between the center pixel and a neighboring pixel is encoded by three values (1, 0 or -1) according to a threshold T. The ternary pattern is divided into two binary patterns taking into account its positive and negative components. The histograms from these components computed over a region are then concatenated. Figure 2 depicts an example of splitting a ternary code into positive and negative codes. It is worth noting that LTP resembles the texture spectrum operator [30], which also used a three-valued output instead of two.

In [30], a completed modeling of the LBP operator was proposed and an associated completed LBP (CLBP) scheme was developed for texture classification and analysis. The image local differences are decomposed into two complementary components: the signs and the magnitudes and two operators, CLBP-Sign (CLBP S which is equivalent to LBP) and CLBP-Magnitude (CLBP M) were proposed to encode them. The center pixels representing the image gray level are also converted into a binary code (CLBP C) by global thresholding. Compared to LBP, the CLBP M and CLBP C were combined with CLBP S as complementary information to improve the texture classification.

In [10], LBP variance (LBPV) was proposed as a rotation invariant descriptor exploiting the contrast information by:

- Putting the local contrast information into one-dimensional LBP histogram. The variance $VAR_{P,R}$ was used as an adaptive weight to adjust the contribution of the LBP code in histogram calculation.
- Learning the principal directions. The extracted LBPV features are used to estimate the principal orientations, and then the features are aligned to the principal orientations;
- Determining the non-dominant patterns and thus by reducing them, feature dimension reduction was achieved.

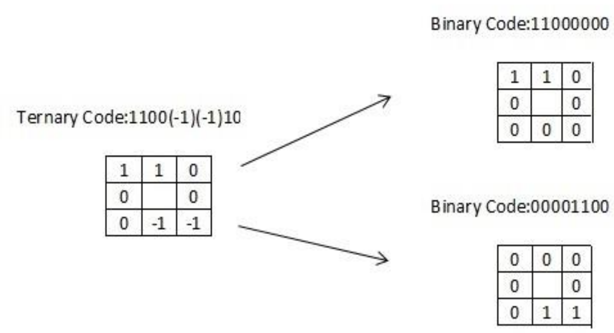

Figure 2: Local ternary pattern operator.

Liao et al. [14] introduced dominant local binary patterns (DLBP) which make use of the most frequently occurred patterns of LBP to improve the recognition accuracy compared to the original uniform patterns. The method has also rotation invariant characteristics. Unlike basic LBP approach which only exploits the uniform patterns, DLBP computes the occurrence frequencies of all rotation invariant patterns defined in the LBP. These patterns are then sorted out in a descending order. The most frequently occurring patterns are expected to contain dominant patterns in the image and, therefore, are selected. It is empirically shown that the DLBP approach can be more reliable than LBP to represent the dominant information in textured images.

The extraction of LBP features is usually done in a circular or square neighborhood. A circular neighborhood is important especially for rotation-invariant operators. However, in some applications, such as face recognition, rotation invariance is not required,but anisotropic information may be important. To exploit this, an elliptical neighborhood definition, calling their LBP variant an elliptical local binary pattern (ELBP). ELBP combined with a local gradient (contrast) measure, provided improved results in face recognition experiments. In another related work, [17] investigated the use of different neighborhood topologies and encodings in their research on LBP variants for medical image analysis.

V. CONCLUSION

General visual features most widely used in content-based image retrieval are color, texture, shape, and spatial information. Although content-based retrieval provides an intelligent and automatic solution for efficient searching of images, the majority of current techniques are based on low level features or current techniques are primarily based on low level features. In general, each of these low level features tends to capture only one aspect of an image property. Neither a single feature nor a combination of multiple features has explicit semantic meaning. LBP captures small and fine details, while Gabor filters encode appearance information over a broader range of scales.

REFERENCES

[1] Dr. Fuhui Long, Dr. Hongjiang Zhang and Prof. David Dagan Feng, “Fundamentals of Content Based Image Retrieval”

[2] A. E. Cawkill, "The British Library's Picture Research Projects: Image, Word, and Retrieval,"Advanced Imaging,Vol.8, No.10, pp.38-40, October 1993.

[3] J. Dowe, "Content-based retrieval in multimedia imaging," In Proc. SPIE Storage and Retrieval for Image and Video Database, 1993.

[4] C. Faloutsos et al, "Efficient and effective querying by image content," Journal of intelligent information systems, Vol.3, pp.231-262, 1994.

[5] Y. Gong, H. J. Zhang, and T. C. Chua, "An image database system with content capturing and fast image indexing abilities", Proc. IEEE International Conference on Multimedia Computing and Systems, Boston, pp.121-130, 14-19 May 1994.

[6] R. Jain, Proc. US NSF Workshop Visual Information Management Systems, 1992.

[7] R. Jain, A. Pentland, and D. Petkovic, Workshop Report: NSF-ARPA Workshop on Visual Information Management Systems, Cambridge, Mass, USA, June 1995.

[8] T. Ahonen, A. Hadid, M. Pietikäinen, Face description with local binary patterns:application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12)(2006) 2037–2041.

[9] T. Ahonen, M. Pietikäinen, Image description using joint distribution of filter bankresponses., Pattern Recognit. Lett. 30 (4) (2009) 368–376.

[10] Z. Guo, L. Zhang, D. Zhang, Rotation invariant texture classification using lbp variance (lbpv) with global matching, Pattern Recognit. 43 (3) (2010) 706–719.

[11] A. Hadid, M. Pietikäinen, T. Ahonen, A discriminative feature space for detecting and recognizing faces, in: IEEE Conference on Computer Vision and Pattern Recognition, vol. II, 2004, pp. 797–804.

[12] A. Hadid, G. Zhao, T. Ahonen, M. Pietikäinen, Face analysis using local binary patterns, in: M. Mirmehdi, X. Xie, J. Suri (Eds.), Handbook of Texture Analysis, Imperial College Press, 2008, pp. 347–373.

[13] D. Huang, C. Shan, M. Ardabilian, Y. Wang, L. Chen, Local binary patterns and its application to facial image analysis: a survey, IEEE Trans. Syst. Man Cybernet. Part C: Appl. Rev. 41 (6) (2011) 765–781.

[14] S. Liao, M.W.K. Law, A.C.S. Chung, Dominant local binary patterns for texture classification, Trans. Image Proc. 18 (5) (2009) 1107–1118.

[15] B. Ma, W. Zhang, S. Shan, X. Chen, W. Gao, Robust head pose estimation using LGBP, in: Proceedings of the 18th International Conference on Pattern Recognition, in: ICPR '06, vol. 02, 2006, pp. 512–515.

[16] T. Mäenpää, M. Pietikäinen, Texture analysis with local binary patterns, in: C. Chen, P. Wang (Eds.), Handbook of Pattern Recognition L. Wolf, T. Hassner, Y. Taigman, Y.: descriptor based methods in the wild, in: Faces in Real-Life Images Workshop in ECCV. (2008), 2008, pp. 1–14.

[17] L. Nanni, A. Lumini, S. Brahnam, Local binary patterns variants as texture descriptors for medical image analysis, Artif. Intell. Med. 49 (2) (2010) 117–125.

[18] C. Ng, Y. Tay, B.-M. Goi, Recognizing human gender in computer vision: A survey, in: PRICAI 2012: Trends in Artificial Intelligence, in: Lecture Notes in Computer Science, vol. 7458, 2012, pp. 335–346.

[19] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal.Mach. Intell. 24 (7) (2002) 971–987.

[20] M. Pietikäinen, A. Hadid, G. Zhao, T. Ahonen, Computer Vision Using Local Binary Patterns, Springer, 2011.

[21] C. Shan, S. Gong, P.W. McOwan, Facial expression recognition based on local binary patterns: a comprehensive study, Image Vision Comput. 27 (6) (2009) 803–816.

[22] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, in: Proceedings of the 3rd International Conference,AMFG'07, 2007, pp. 168–182.

[23] L. Wolf, T. Hassner, Y. Taigman, Y.: descriptor based methods in the wild, in: Faces in Real-Life Images Workshop in ECCV. (2008), 2008, pp. 1–14.

[24] Z. Yang, H. Ai, Demographic classification with local binary patterns, in: The 2nd International Conference on Biometrics, 2007, pp. 464–473.

[25] Z. Yang, H. Ai, Demographic classification with local binary patterns, in: The 2nd International Conference on Biometrics, 2007, pp. 464–473.

[26] I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[27] K. Elissa, "Title of paper if known," unpublished.

[28] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[29] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[30] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

