Integrating SIFT Features with Multivariate Gaussian Distribution to Enhance CBIR Performance

Vamsidhar Yendapalli[1], Ramesh Naik Bhukya[2], Mylara Reddy C[3]
[1]Professor, CSE, GITAM School of Technology, Bengaluru, India, vdhar@gitam.edu
[2]Asst.Prof, CSE, GITAM School of Technology, Bengaluru, India, rameshnaik3@gmail.com
[3]Asst.Prof, CSE, GITAM School of Technology, Bengaluru, India, mchinnai@gitam.edu

ABSTRACT

CBIR gained substantial importance in the recent fifteen years. Variety of modern systems have designed to effectively retrieve images from database and recognition of the objects in computer vision. Among all methods, Histogram of Gaussian (HOG) and Scale Invariant Feature Transform (SIFT) are standard procedures for attaining the features of the image. The image features achieved by SIFT are helpful in several applications, but these features prone to zero-order statistics as they derived from histogram quantities. This results in; there lacks the natural mechanism of high descriptiveness of image features. A novel method is developed based on Multivariate Gaussian distribution which establishes the relation among SIFT features in the neighbourhood. The exciting part of this approach lies in representing the non-linear Gaussian space into linear space as the Gaussian space is in Riemannian manifold. The Gaussian space can be mapped into linear Euclidean space using Lie group theory. The tests were conducted systematically on Caltech-101 and ZANG image databases to endorse this approach. The experiments demonstrated that a significant amount of the CBIR system performance has improvised with the proposed approach.

Key words: Computer vision, Descriptors, Image retrieval, Multi-variate Gaussian distribution.

1. INTRODUCTION

Content-Based Image Retrieval (CBIR) achieved significant impact in recent decades in which the similar images extracted to query model based on a low level and derived visual features. Extensive exploration has absorbed to distinguish local characteristics of the image in the past ten years. Various descriptors have invented for demonstrating the area of interest in the pictures confined by region detectors, colour gradients detectors and shape detectors [1], [2]. Extraction of local characteristics of the image plays an ultimate role in the success of many image descriptors in the vision tasks. The complicated job in CBIR is to shape up features which are high uniqueness, i.e. features are invariant to geometrical transformations. Belongie et al. [3], [7] proposed a texture-based transformation approach to represent image area which is articulate in and texture space using multi-orientation filter banks and the second-moment matrix. They additionally utilized the Gaussian window and Expectation-Maximization (EM) algorithm to achieve differentiation and scale invariance. Zhang et al. [4], [8] developed the CBIR

system using Gabor texture features for image retrieval. Texture features obtained by the mean and variance of the images which are filtered Gabor approach. Variation standardization in their approach is comprehended by a circular shift of the extracted image values so that all images have the same dominant direction. Hiremath and Jagadeesh[5] presented a novel process for relating texture, colour, and shape information and attained higher retrieval efficiency. In their method, the image segregated into non-intersecting tiles of equivalent size. The local descriptors of an image obtained based colour and texture attained by utilizing moments as well as Gabor filtered features of these tiles, respectively [1]. Roman et al. [7] demonstrated a machine learning system for obtaining low-dimensional image features in higher-level Gaussian Process (GP) tasks. Such issues are progressively numerous and indispensable but have previously revealed severe practical implications. To overcome this issue, they alternatively present a novel algorithm for marginalizing GP hyperparameters. This approach would give marginal predictions and better performance than previous hyperparameter correction. Venugopal et al. [6] projected a novel technique for shape-based image extraction using Adapted Fourier Descriptors which obtain shape information of an image that is distinct to rotation, scale, translation.

Peihua et al. [8], [9] proposed an image classification technique using multivariate Gaussian distribution which also describes the process of transforming Gaussian into linear space by specifying a multiplication operation on the manifold. Dongsheng et al. [10] described a Generic Fourier descriptor (GFD) to diminish the limitations of prevailing shape representation techniques. Their proposed shape descriptor originated by sampling 2-D Fourier transform on a polar shape image. The attained shape descriptor is robust and independent concerning applications. Xiangyuan Zhao and Brian Nutter [11] presented a CBIR system by utilizing Deep Autoencoder (DAE) and Wavelet Transformation. Wavelets transform are a promising method for image retrieval based on texture properties. Montazer et al. [12] introduced a synchronized methodology utilizing SIFT and wavelet transform to define the picture characteristics effectively. In CBIR, as massive image information bases are being gathered, sorting out the image data sets for effective retrieval is a non-trivial issue. A classic function valued descriptor is innovatively exceptional, which separates visual picture data and utilized in the few application of computer vision, such as object movement recognition, perceiving objects, object classification [18].

Generally, visual information exploited using local descriptors such as SIFT or HoG [20], [24]. The image data acquired by applying SIFT technique is productive in many

image processing applications. However, it just acquires low-level features such as histogram frequencies of an image, i.e., and there is no association established between the image features and this results in accurate definiteness of image features [12], [20]. Too, histograms frequencies generally experience quantization issue. Multivariate Gaussian (μ, $\sum$) is a noticeable methodology to originate proficient descriptiveness of image information in higher dimensional space N(d) [8], [16]. The foremost disadvantage of this methodology is Gaussian space prone to non-linear space, i.e. non-linear space defined by Gaussian usually lies in a manifold. The non-linear Gaussian space can be transformed into Linear Euclidean space by characterizing a direct logarithmic procedure on manifold [8], [15], [21] to resolve this issue. The technique for transforming of Gaussian space into linear space (Euclidean space) has exploited in detail in this paper. The subsequent features have distinguished to as Log-Euclidean Multivariate Gaussian descriptors.

The rest of this paper has composed as follows. Section 2 demonstrates the specifics of SIFT descriptors. Section 3 presents the Riemannian manifold. In Section 4, Gaussian Embedding in Linear Space has presented. Section 5 describes Log Euclidean Multivariate Gaussian Descriptors. Section 6 explore the test result to estimate and analyze our descriptors. Finally, Section 7 concludes the paper.

2. RELATED WORK

2.1 Introduction SIFT

The essential aspect of CBIR in pattern analysis and computer vision industries is to extract correlated images from an extensive image store efficiently. However, the specific aspect of this methodology is the accurate extraction of images because images are often prone to geometrical changes. i.e. while capturing an image, there could be rotational variance, changes in posture, disorderly background, scale changes, brightness variations and incomplete occlusion [9], [10], [12]. Determining of low-level perceptual features using SIFT is the prospective methodology in which object features in the images selected based on visual pixel quantities which can then distinguished to other features of an image [20], [22], [35]. In this methodology, the set of arguments can be used to extract visual information; but then how many arguments and all arguments are necessary or optional is not comprehended. The organization of default arguments is portrayed in the first paper by David Lowe [12], [20], yet whether these all are necessary or optional is not clear.

The essential background of the SIFT procedure has introduced in this section. The SIFT approach has comprehended in four-stage process that obtain accurate useful characteristics from an image which are extremely reasonable for the vision applications such as image tracking, recognition and categorize objects [12]. These characteristics considered to be free from rotational changes, scale variations, noise, small changes in viewpoint and, illumination variance. Clusters of features from an image that is similar to a cluster of characteristics from another image may indicate, with a high likelihood, areas that match. The four stages of the SIFT algorithm is given in Lowe's paper [20]. The SIFT procedure categorized into four stages, as following.

2.1.1 Scale-space Extrema Detection.

Finding the scale at which the images features are more discriminate is determined in the scale space extrema detection. Multiple octaves are created from blurred pictures to identify scale-invariant areas through dissimilar scales using the convolution of Gaussian, as specified in Eq. (1). The variance of Gaussian designated as the change between two consecutive images inside the identical octave as shown in Eq. (2) [45]. Let the input image represented as I(x, y) and the scale space determined by convolution with Gaussian kernel :

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \qquad \text{Eq. (1)}$$

In the Eq. (1), * is the convolution operator, and the function computes G (x, y, σ): The difference of Gaussian DoG(x, y, σ) among two images has calculated, first with L(x, y, σ) and next with k times L(x, y, σ). DoG(x, y, σ) is then shown in the Eq. (2):

$$DoG(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \qquad \text{Eq.(2)}$$

Local minima and maxima can be determined by comparing eight neighbours at an identical scale, and its nine neighbours up and down one scale from DoG. Minima or maxima of all successive 26 points has called to as a local extremum.

2.1.2 Feature Localization

The characteristics which are having minimal effect in scale space have disposed of in this stage. DoG function normally indicates high reaction at edges in the fact that area along the border is ineffectively defined and consequently sensitive to the minimal effect of distortion in the image. Interpolation is a beneficial procedure to remove undesirable edges and pixels which are in the neighborhood of low illumination.

i. Local Extrema Detection:

Neighbours of each keypoint equated to distinguish local maximum and minimum of DoG images. Keypoint around regions 3x3 of DoG at neighboring and present scale are examined to recognize local maximum and minimum.

ii. Rejecting Low Contrast Keypoints

The low contrast points can be eliminated because these points are susceptible to noise or are poorly determined along edges. A quadratic function of three dimensions applied to the local example points to govern the position of the maxima. A Taylor series is expanded (up to the quadratic terms) on to the DoGD(x; y; σ), and Origin of low Contrast points can be determined shown in Eq. (3), so the origin is located at the sample point.

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} \text{Eq. (3)}$$

In Eq. (3), x is offset from a particular point, D as well as its derivatives estimated at each sample point. The region of the extremum $\hat{x}$ is obtained by calculating the derivative of D concerning offset x=(x; y; σ)T and taking it to zero as shown in Eq. (4).

$$\hat{x} = \frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x} \qquad \text{Eq. (4)}$$

If the value of offset $\hat{x}$ is greater than 0.5 in any dimension implies that extremum point lies nearer to any other sample point.

iii. Eliminating Edge Responses:

The edges of the DoG disclose great reaction at the peak, pitiable other direction. 2x2 Hessian matrix (H) used to calculate basic curvature. To detect eigenvalues of the boundaries, Harris corner detector is a utilized in which

eigenvalue of one edge dissimilar from others. The Hessian matrix (H) is given as

$$H = \begin{vmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{vmatrix}$$

Tr(H) is a trace, Det(H) is determining factor of H resulting from Eq. (5), (6) respectably. Let α be the eigenvalue with higher degree and β the smaller.

$$Tr(H) = D_{xx} + D_{yy} = \alpha + \beta \quad \text{Eq. (5)}$$
$$Det(H) = D_{xx}D_{yy} - (D_{xy})^2 \quad \text{Eq. (6)}$$
$$\frac{Tr(H)^2}{\det(H)} = \frac{(\alpha+\beta)^2}{\alpha*\beta} = \frac{(\alpha\beta+\beta)^2}{\alpha\beta^2} = \frac{(r+1)^2}{r} \text{Eq. (7)}$$

If ∞ and ß are equally large, a curve is originated.

If ∞<< ß or ∞ >> ß an edge is originated.

If ∞ and ß taking smallest values, there are no key points of attention at this pixel (x; y). Thus, the expansion of corner feedback can be calculated by r (=r ß). (r+1)2/r is at a min when the two Eigenvalues have equally derived from Eq. (7).

2.1.3 Orientation Assignment

Typically, pictures have captured at various angles; thus obtained features might be inclined to geometrical changes. Rotational invariance can be accomplished by appointing each point is gradient slope in its locality. Slope has designed with 32 bin direction histogram, and noteworthy keypoints closer to the middle have more effect on direction. Each key direction is one different component [10].

Let L be the Gaussian flattened image and *m* is gradient magnitude which is taken from scale space, the computation of orientation is given in Eq. (8):

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \text{Eq. (8)}$$

Orientation (θ) is

$$\theta(x,y) = \tan\left(L(x,y+a)\frac{L(x,y-1)}{(L(x+1,y)-L(x-1,y))}\right)$$

For each key point form histogram of orientation from Gaussian smoothed image.

Highest peak and variable peaks below 80% of the extreme of the histogram have used to make a key points. A portion of the focuses can be appointed different directions. Uppermost peaks and alternative peaks at intervals 80% of extreme of the histogram are taken to produce key point. Barely any more number of the keypoints might be allocated different directions.

2.1.4 Creating the Feature Descriptor

A Descriptor derived from an image made with 128-dimensional array which represent pixel characteristics of region surrounding the key point and has appeared in Figure 1. A 4x4 size histograms computed, which is positioned on the key point and rotated to organize the key orientation in the all directions. The procedure of producing feature descriptors have specified below [42].

i. Divide the keypoint segment into 4x4 sub-segments.

ii. Generate histogram for each sub-region, 16 bins utilized to create 16 histograms. Trilinear Interpolation has utilized to place features into histogram bins. Tests are Gaussian weighted with N(σ) = 1.5 district width.
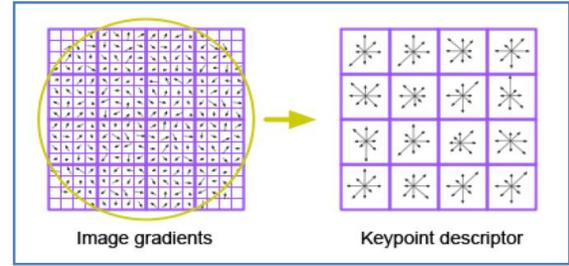

Figure 1: Image gradients and key point descriptors

To generate keypoint descriptors [29], [34], [47] the native gradient statistics also utilized. Interchange the gradient features to contrast with the keypoint orientation then Gaussian function could be utilized to reweight the occurrences with N(σ) = 1.5 keypoint scale. With this data, histogram at given point made over a window focused on the key point. This resulting feature vector is consisting of 128 elements.

3. INTRODUCTION TO RIEMANNIAN GEOMETRY

Let M be a manifold that could be a mathematical space locally seems like Euclidean space with all of its usual topologies. A topological space could be a set of points together with its neighbourhoods for every pixel. In other words, the features a neighbourhood that's homeomorphism to Euclidean space of n-dimension [15], [26]. Homeomorphism could be a continuous perform that maps one mathematical space to a different space and features a continuous mathematical function. Mathematically dimensions are degrees of freedom. A point is zero-dimensional said to be no degree of freedom. The line is one-dimensional so require x-direction, a plane is two dimensional hence need both the x and y directions, a sphere is three-dimensional requires x, y, and z directions. We can even extend that to hyperspheres, which are four-dimensional, or to something more intuitive like the four dimensions of space-time to an infinite number of sizes. If there exists a homeomorphism between the objects, then they said to be homeomorphic.

A Riemannian manifold has constructed from the three-layer topological structure. Notions of Topology, for example, Continuity and merging are characterized in the topological layer. Stretching out the notions of differentiability to the manifold is permitted in differential layer and the inflexible geometrical magnitudes, for example, space, edges and arch on the manifold are permitted in Riemannian manifold. A Riemannian manifold ($\mathcal{M}$, g) has furnished with a Riemannian metric g stretched out from differential manifold [24]. Neighbourhood inner products determine the Riemannian metric g on tangent vectors. i.e

$$g_x(\,.\,,\,.\,) : T_x\mathcal{M} \times T_x\mathcal{M} \to R, x \in \mathcal{M}$$

The measure g of Riemannian manifold is bi-linear, symmetric, positive definite and is also C^∞ differentiable in x.

$$g_x\left(\sum_{i=1}^{n} u_i, \sum_{i=1}^{n} v_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} g_x(u_i, v_j)$$
$$g_x(u,v) = g_x(v,u)$$
$$g_x(u,u) \geq 0$$
$$g_x(u,u) \geq 0 \Leftrightarrow u = 0$$

The measure g is determined $\forall$ u, v $\in$Tx $\mathcal{M}$, based on bi-linearity of the inner product as in Eq. (14).

$$g_x(v,u) = \sum_{i=1}^{n}\sum_{j=1}^{n} v_i u_j g_x(\partial i, \partial j) \quad \text{Eq. (14)}$$

The absolute meaning of gx in Eq. (14) have acquired by $\{gx(\partial i, \partial j) : 1 \le i, j \le n\}$– the arrangement of internal items between the basic components $\{\partial i\}_{i=1}^{n}$ of Tx $\mathcal{M}$. Where $g_x(\partial i, \partial j)$ is Gram matrix $[G(x)]ij$ and is an asymmetric and positive clear grid that portrays the metric g_x [17]. Thus the lengths of tangent vectors v $\in$Tx $\mathcal{M}$ is portrayed by $\sqrt{g_x(v,u)}$ and lengths of curves γ: [a, b] $\rightarrow \mathcal{M}$ is characterized by

$$L(\gamma) = \int_a^b \sqrt{g_x(\dot{\gamma}(t), \dot{\gamma}(t))}dt \quad \text{Eq. (15)}$$

In Eq. (15), $\gamma\dot{}$ (t) is the speed vector of the curve γ at time t. The separation dg(x, y) is geodesic among two points x, y $\in \mathcal{M}$ is characterized from Eq. (15) by Eq. (16).

$$d_g(x,y) = \inf_{\gamma \in \Gamma(x,y)} \int_a^b \sqrt{g_x(\dot{\gamma}(t), \dot{\gamma}(t))}dt \text{Eq. (16)}$$

The interfacing x and y points of curves have characterized by Γ(x,y). The geodesic curves depicted by geodesic distance dg and negligible curve accomplishing it. Geodesic space fulfils the typical necessities of space, with the topological structure. On the off chance that the manifold being referred to is obvious from the unique circumstance, we will expel the subscript and utilize d for the geodesic space. Specifically, smaller manifolds are geodesically complete if any geodesic curve c(t), t $\in$ [a, b], can be stretched out to be characterized for all t $\in$ R.It can be shown that the following are equivalent.

• ($\mathcal{M}$, g) is geodesically complete
• d_g is a complete metric on $\mathcal{M}$
• closed and bounded subsets of $\mathcal{M}$ are compact.

The Hopf-Rinow theorem [44] focusses that a geodesic can joint any two points if $\mathcal{M}$ is geodesically complete. Assuming that Riemannian manifolds ($\mathcal{M}$, g), ($\mathcal{N}$, h) and a isomorphism of smooth manifold samongst them f: $\mathcal{M} \rightarrow \mathcal{N}$. We characterize the pull-back and push-forward maps below.

Definition 1. The push-forward map f*: Tx $\mathcal{M} \rightarrow$ Tf(x) $\mathcal{N}$, associated with the diffeomorphism f: $\mathcal{M} \rightarrow \mathcal{N}$ is the mapping that satisfies [19], [38]:

v (r o f) = (f * v)r, $\forall$r $\in$C$^\infty$ ($\mathcal{N}$,R).

The push-forward has described from the coordinate-free version of the Jacobian Matrix J or the total derivative operator associated with the local chart representation of f. In other words, if we define the coordinate version of f: $\mathcal{M} \rightarrow \mathcal{N}$ [39]

$$\tilde{f} = \emptyset \, ofo\psi^{-1}: R^n \dashrightarrow R^m$$

Where $\emptyset$ and ψ are local charts of N, $\mathcal{M}$ then push forward map is :

$$f * u = Ju = \sum_i \left(\sum_j \frac{\partial \tilde{f}\iota}{\partial x_j} u_j\right) e_i$$

where J is the Jacobian of $\tilde{f}$ and $\widetilde{f}\iota$ is the i-component function of $\tilde{f}$: Rm $\rightarrow$ Rn. Intuitively, as illustrated in Fig 2, the push-forward transforms velocity vectors of curves γ to velocity vectors of transformed curves f(γ).

Definition 2. Given (N, h) and a diffeomorphism f : $\mathcal{M} \rightarrow \mathcal{N}$ we define a metric f * h on $\mathcal{M}$ called the pull-back metric by the relation (f * h) x (u, v) = h f(x) (f * u, f * v).

Definition 3. An isometry is a diffeomorphism f: $\mathcal{M} \rightarrow \mathcal{N}$ between two Riemannian manifolds ($\mathcal{M}$, g), ($\mathcal{N}$, h) for which

gx(u, v) = (f * h) x (u, v) $\forall$x $\in \mathcal{M}$, $\forall$u, v $\in$Tx $\mathcal{M}$.

Two Riemannian manifolds identified as identical from the above said isometries in terms of their Riemannian structure. Consequently, isometries assure all the geometric properties including the geodesic distance function dg(x, y) = dh(f(x), f(y)). Note that the above definition of an isometry is defined through the local metric in contrast to the global definition of isometry. A Riemannian structure and smooth manifold may be defined over a topological manifold with a boundary as well. The definition is a certainly straightforward expansion using the notion of differentiability of maps between non-open sets in R^n [40].

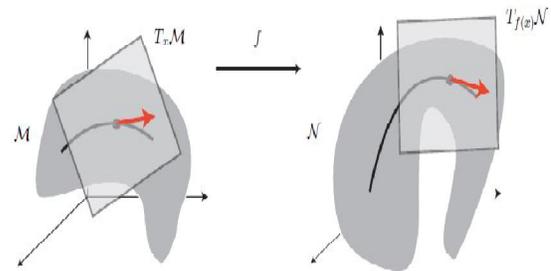**Figure 2**: The map f: $\mathcal{M} \rightarrow \mathcal{N}$ describes a push forward map f*: Tx $\mathcal{M} \rightarrow$ Tf(x) $\mathcal{N}$ that converts velocity vectors of arcs to velocity vectors of the converted arcs.

3.1 Lie Group Algebra

Let a group G which has furnished with a multiplication operation, i.e. G $\times$ G $\rightarrow$ G, yield a component which intersect any two components *a* and *b* in G to written *a · b*, of G fulfils the given characteristics [32], [36]:

(i) For every element *a, b* and *c* in G, *(a · b) · c = a · (b · c)* is associative.

(ii) For every *a* in G, *a · e = e · a = a*. Where e-identity component.

(iii) There is an inverse a-1 for every component *a* in G, for which *a · a-1 = a-1 · a = e*.

(iv) For all a, b in G, if *a · b = b · a*, then G is said to be a commutative (or abelian) group.

A subsection of Group G is F, if this subsection produces a group under Multiplication activity, then F said to be a subgroup of G [36], [37]. The group multiplication and opposite are harmonious activities in the Lie Group theory. A Lie group is a smooth activity that observe manifold and fulfils further circumstances that are differentiable on Gaussian space. Accordingly, the locality of any element can be reasonably represented by its Euclidean space [32]. Let •, ∘ be the multiplicative activities on be Lie samples GL, GL′ correspondingly. A Lie group is assumed to homomorphic (φ : GL $\rightarrow$ GL′) and a smooth transitionthat satisfies $\varphi(a • b) = \varphi(a) ∘ \varphi(b)$ for everyone of *a, b*$\in$ GL.

3.2 Matrix Group

The Square Matrices are imperative cases to progress Lie Algebra. Let A be the regular of nxn matrices which

frames a Lie Group if A is invertible under matrix multiplication at that point mentioned to as a general linear group. The Matrix groups, by and large, got from Lie subgroups of the general linear groups. Let a square matrix A and exponential of A, indicated by exp(A). The arrangement mapping is smooth or any X and exp(X) [16], [46]. Let the matrix B, and Log on B is meant by log(B), is a matrix A to such an extent that exp(A) = B. Let the arrangement of every single real number R and its Lie Algebra R+. The Log-Euclidean notation is detailed as follows.

$exp: R \rightarrow R^+$, x → exp(x) The exponential of real numbers is a smooth bijection, and its inverse log is also smooth.

This substantiate exponential bijection *exp* and its inverse log is a both diffeomorphism. we characterize

$\otimes: R^+ X\ R^+ \rightarrow\ R^+, x \otimes y = exp\ (log(x) + log(y))$ Eq. (17)

$\odot: R\ X\ R^+ \rightarrow R^+, \lambda \odot x = exp\ (\lambda\ log(x)) = x^\lambda$ Eq. (18)

The Eq. (18) articulates employing logarithm, the multiplications in the Lie group R^+ are mapped to the additions in the logarithm area. This is composed to as "Log-Euclidean".

4. LOG-EUCLIDEAN MULTIVARIATE GAUSSIAN DESCRIPTORS

Let us consider M a given input image, and first, the characteristics of M are obtained by utilizing an importance point detector like SIFT. The process of obtaining SIFT characteristics is described in section II. Let X= be the basic features obtained for an image in the database by utilizing SIFT algorithm [11], [34] i.e.

$$X = \begin{bmatrix} a_{11}a_{12}a_{13} \dots \dots a_{1\ 127}a_{1\ 128} \\ a_{21}a_{22}a_{23} \dots \dots a_{2\ 127}a_{2\ 128} \\ \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \\ \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \\ a_{n1}a_{n2}a_{n3} \dots \dots a_{n\ 127}a_{n\ 128} \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ . \\ . \\ \mu_n \end{bmatrix}$$

The SIFT features are effective in a wide-range of systems. Still, these features exhibit low order statistics as they only derived based on histogram frequencies and as a result there lacks high discriminative properties of images. Besides, distinct histograms typically suffer from the quantization issue [8], [41]. Multivariate Gaussian (μ, ∑) is prominent approach for effectively discriminate image characteristics in d-dimensional space N(d) [8], [16]. Image retrieval framework has shown in figure 3. The characteristics for all images are attained by utilizing our described method and are placed in the feature database. For a user input image, the characteristics are attained and then these characteristics have contrasted with all the image characteristics placed in the feature database. The obtained outcomes are indexed and then given to the end user. The distribution of Multivariate Gaussian [16] of a regular d-dimensional vectors X is obtained by Eq. (13) i.e.

variable Gaussian (μ , $\sum$) may be represent image options in d-dimensional house N(d) [8], [16]. Image retrieval framework has shown in figure 3. For all pictures within the image database, the features are extracted by victimization our planned approach and are keep within the Feature database. For a given question image, the options are extracted and so these features have compared with all the image options keep within the feature database. The retrieval results are indexed and so show to the user. The

distribution of variable Gaussian [16] of a collection of d-dimensional vectors X is calculated by Eq. (13) i.e.

$$f(X|\mu_i, K) = (2\pi)^{-d/2}(detK)^{1/2}e^{-(x-\mu_i)^T K^{-1}(x-\mu_i)/2}\text{Eq. (19)}$$

Where μ is the mean vector, det is the basis, and K is the covariance space with real symmetric affirmative semi-definite matrices [17] defined as follows:

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} a_{ij}$$

$$K = \begin{bmatrix} S_{11}S_{12}S_{13} \dots \dots S_{1\ 127}S_{1\ 128} \\ S_{21}S_{22}S_{23} \dots \dots S_{2\ 127}S_{2\ 128} \\ \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \\ \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \\ S_{n1}S_{n2}S_{n3} \dots \dots S_{n\ 127}S_{n\ 128} \end{bmatrix}$$

Where $S_{ii} = \frac{1}{N-1}\sum_{j=1}^{n}[a_{ij} - \bar{\mu}_i]^2$

$$S_{jk} = \frac{1}{N-1}\sum_{i=1}^{n}[a_{ij}a_{ik} - n\bar{\mu}_j\bar{\mu}_k]$$
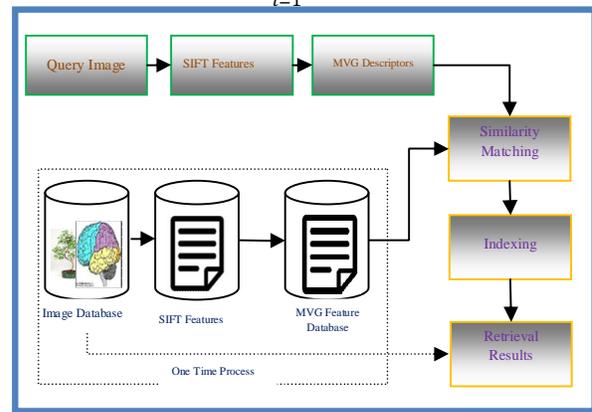

Figure 3: Image retrieval framework

The way toward mapping Riemannian complex to Euclidean space has portrayed as follows:

To start with, the covariance matrix projected on vector space tangent to the Riemannian manifold.

$t_K = log_p(K) = Q^{\frac{1}{2}}log\ (Q^{-\frac{1}{2}}KQ^{\frac{1}{2}})Q^{\frac{1}{2}}$ Eq. (20)

In Eq. (20) log(Q) is Manifold particular logarithm operator and the log is Matrix logarithm operator.

At that point, the orthogonal directions of the anticipated vectors are extracted as in Eq. (21).

$vec_p(t_K) = vec_I\left(Q^{-\frac{1}{2}}t_KQ^{-\frac{1}{2}}\right)$ Eq. (21)

While the vector operator on the tangent space at the character of a symmetric matrix Y is characterized as:

$vec_I(Y) = \left[y_{1,1}\sqrt{2}\ y_{1,2}\sqrt{2}y_{1,3}y_{2,2}\sqrt{2}y_{2,3}\dots\dots y_{d,d}\right]$ Eq. (22)

Replace with tK in Eq. 22, the projection of K on the hyperplane tangent to Q progresses toward K.

$$K = vec_I\left(log\left(Q^{-\frac{1}{2}}KQ^{-\frac{1}{2}}\right)\right)\ \text{Eq. (23)}$$

The projection point Q could impact the execution (mutilation) of the projection since it is irregular in nature. The Identity matrix is utilized, which essentially interprets standard matrix logarithm. So forward a (d2 + d)/2-dimensional element vector is acquired since K is the symmetric framework. The covariance matrix which is gotten from SIFT descriptors is anticipated on a Euclidean

space and linked to the mean vector to acquire the final descriptor. The exploratory outcomes demonstrate that the resultant descriptor esteems low, while few are high. The power standardization techniques can be adjusted to disperse the values more equally.

5. RESULTS AND ANALYSIS

We complete tests on the Coral (Wang) Database and Caltech - 101database [30] to make an investigation of parameters engaged with LEMVG. The Coral (Wang) image Database contains 10,000 pictures altogether and isolated into ten classes. We take after the standard measure Average Precision (AP) over various classes of the photographs [31].

Caltech-101 has around 9K images conveyed in 101 categories, comprising of multiple poses, sizes and at different lighting conditions. The depiction of the Caltech-101 database is appeared in fig-4 [30]. The target of Log Euclidean Multivariate Gaussian Descriptors (LEMVG) is to depict local features of the pictures efficiently. The often utilized basic features, for example, intensity, colour, location, first and second derivatives are figured out at various scales. We applied distinctive operators to represent basic features extracted in diverse directions and at multiple scales. Table -1 presents the image retrieval results obtained from numerous combinations of operators over WANG Database, and Table -2 gives the image retrieval results obtained from a various variety of operators over Caltech-101 Database as shown in Fig-4.

In the Table-1&2, 1st row presents image retrieval results which compute covariance descriptors [9] based on combination essential features intensity, location [18]. In the Table-1&2, 2nd row shows results based on features calculated by using orientation histogram of edges (OHE) [24], [31]. OHE usually collects the zero-order statistics of the image. In Table -1&2, 3rd row presents results based on the basic features computed by the first-order derivative operator of Gaussian (FDOG). Finally, in Table -1&2, 4th row shows results based on the basic features computed by evaluating additional colour, Gabor filters.
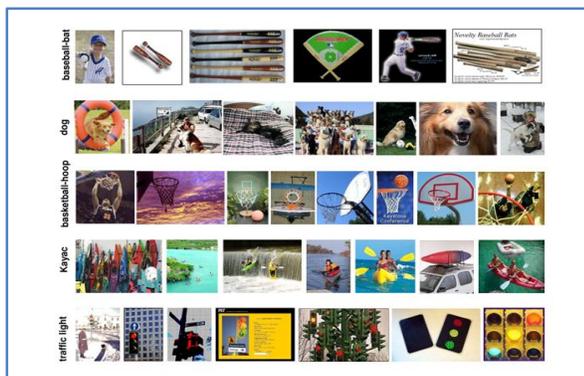

Figure 4: Snapshot of Caltech 101 Image Database

The Experimental results demonstrated that LEMVG Descriptor achieves the best performance over WANG database with a mix of first and second derivatives. Similarly, LEMVG best performance over the Caltech-101 database by utilizing Covariance Descriptors. The retrieval result @2 and @3 of table-1&2 proved that the OHE and Gabor Filter marginally enhance performance over WANG database.

The Gabor filters are of generally high dimension yet neglect to bring better results. This might be because texture data of little patches extricated by Gabor filters isn't sufficiently particular and is consequently not competing for image retrieval task. Graph-1 introduces the examination of Retrieval results which are acquired utilizing LEMVG technique over WANG and Caltech-101 databases. In the Graph-1 AP precision of 10 images taken into consideration at 10% the recall is 80 and at 20% the recall is 76 etc,. The AP on WANG and Caltech-101 databases at 100% Recall is 25 and 19 respectively. We demonstrate that LEMVG has fundamentally the same as performance over both databases, by and large, the results on Wang database outperformed more than 2.4% for numerous scale.

Table 1: The Experimental results of LEMVG obtained using different operator on WANG Database

No	Raw Features	Dimension	LEMVG (AP,%)
@1	Covariance Descriptors (Intencity, Location)	7	53.71
@2	Orientation Histogram of Edges(8 bins)	9	41.72
@3	First-order The derivative operator of Gaussian (FDOG)	12	51.48
@4	Gabor filters	15	50.13

Table 2: The Experimental results of LEMVG obtained using different operator on the Caltech-101 database

No	Raw Features	Dimension	LEMVG(AP,%)
@1	Covariance Descriptors (Intensity, Location)	7	52.44
@2	Orientation Histogram of Edges (8 bins)	9	44.55
@3	First-order The derivative operator of Gaussian (FDOG)	12	47.93
@4	Gabor filters	15	49.83

Further, we did examinations to check our mapping strategies that effectively utilize the geometrical structure of Euclidean space. The performance of direct mapping systems as in eq-19, indirect mapping as in eq-20 and eq-21, are assessed over patch size 12X12. The retrieval results are introduced in Fig-5. The outcomes demonstrated that indirect mapping procedure accomplishes better AP over direct mapping. Out of all direct and indirect mapping outperforms the benchmark method.
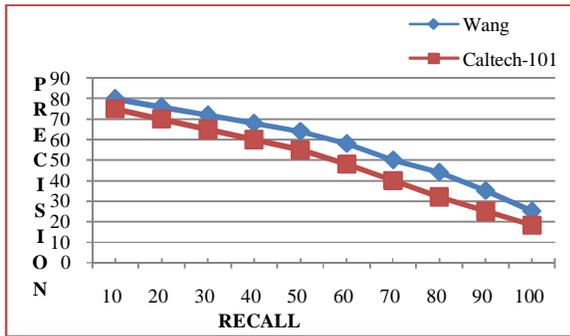
Graph 1: Average Precision of LEMVG Method at a different level of recall over Wang and Caltech-101 databases

The fixed size has an influence on some set of neighborhood key points and local geographies of an image. We chose the training sample step to 2 to examine the influence of patch size, and the results are presented in Table-3. The performance raises consistently as the patch size declines gradually from 24×24 to 12×12. This exhibits that neighborhood characteristics at finer scales are more particular and discriminative. Yet, a too-small patch size (8×8 or littler) prompts a deficient number of samples for Gaussian estimation. By blend of four measures of patches, LEMVG accomplishes AP 53.75%.
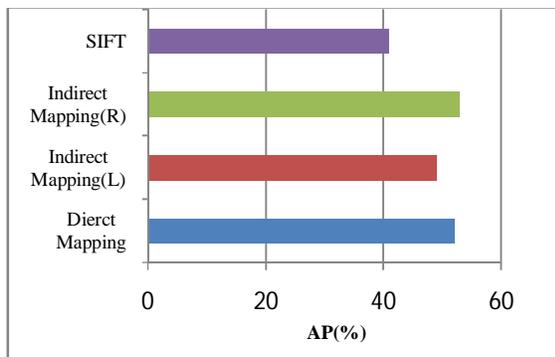

Figure 5: Comparison of Different mapping methods on WANG Database.

The Retrieval outcomes are distinguished with outcomes obtained by SIFT and Linear Discriminant Analysis (LDA) which is expansion of Fisher's Discriminant Exploration over WANG and Caltech-101 databases [33], [16]. LDA is a Machine learning approach which productively separates picture descriptors. The Descriptors utilizing LEMVG and LDA are registered with 12x12 patch size [43], [44]. Table-5 shows the retrieval results Average Precision over Caltech-101 and WANG database, which speaks to their performance against different baseline descriptors.

Table 3: Effect of Patch Size on LEMVG

Patch Size	8X8	12X12	16X16	24X24
WANG	51.32	53.75	51.35	49.30
Caltech-101	49.35	51.45	49.10	47.04

Table 4: Retrieval correctness Precision of SIFT, HoG, LEMVG at different levels of recall

REC ALL	LDA	SIFT	LEM VG
10%	67.85	65.66	75.54
20%	63.54	62.64	70.34
30%	57.75	56.67	67.32
40%	53.45	55.35	63.54
50%	50.32	48.94	54.23
60%	47.45	46.24	50.32
70%	45.56	43.45	47.54
80%	42.12	38.43	42.54
90%	33.12	31.43	39.54
100%	28.12	27.44	35.5
AP	48.92	47.68	54.64

From 101 classes of Caltech-101 image store, ten pictures arbitrarily browsed from every type and are utilized as a query. For each query, typical precision and retrieval at each degree of the recall is generated. The image retrieval result Average Precision of various methods at each degree of recall is given in table-4. In the table-4 precision of LDA, SIFT and LEMVG of Barrel-0004 at 10% the recall is 66.75, 65.56and 75.54 respectively. With Log Euclidean Multivariate Gaussian, we further more realize enhanced precision of 6.56% by and large.

6. CONCLUSION

We have proposed a Log Euclidean Multivariate Descriptors to depict neighborhood, higher-level discriminative features of images. We have developed a Log-Euclidean measurement to transform the Gaussians with Euclidean space instead of Riemannian ones. Not at all like histogram-based descriptors, for instance, SIFT or HoG, the proposed descriptors are steady and models higher-order features of pictures which are remarkably beneficial in Image Retrieval task. We unveiled that the space of Gaussians can be outfitted with a Lie group structure and that it is proportionate to a subgroup of the upper triangular subgroup. These conclusion gives insights into the numerical and geometrical structure of Gaussians. Observational assessment proved the productivity of the subsequent algorithms on both syntactic and real-world problems in image retrieval. With the accomplishment of covariance descriptors, we are additionally intrigued to apply LEMVG descriptors to other vision tasks, for example, object tracking in videos and visual tracking.

REFERENCES

1. K. van de Sande, T. Gevers, and C. Snoek, "**Evaluating color descriptors for object and scene recognition**," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 9, pp. 1582–1596, Sept 2010.
2. Gong, T. Wang, and F. Liu, "**Shape of Gaussians as feature descriptors,**" in Proc. Int. Conf. Comp. Vis. Patt. Recog., 2009, pp.2366–2371.
3. Belongie, Serge, Carson, C., Greenspan, H., & Malik, J. "**Color-and texture-based image segmentation**

using EM and its application to content-based image retrieval." Computer Vision, 1998. Sixth International Conference on. IEEE, 1998.

4. Zhang, Dengsheng, et al. "**Content-based image retrieval using Gabor texture features**." IEEE Transactions PAMI (2000): 13-15.

5. Hiremath, P. S., and Jagadeesh Pujari. "**Content based image retrieval using color, texture and shape features.**" Advanced Computing and Communications, 2007. ADCOM 2007. International Conference on. IEEE, 2007.

6. Venu Gopal, T., B. Ramesh Naik, and V. Kamakshi Prasad. "**Image retrieval using adapted Fourier Descriptors.**" International Journal of Signal and Imaging Systems Engineering 3.3 (2010): 188-194.

7. Garnett, Roman, Michael A. Osborne, and Philipp Hennig. "**Active learning of linear embeddings for Gaussian processes.**" arXiv preprint arXiv:1310.6740(2013).

8. Peihua Li, Qilong Wang, Hui Zeng, Lei Zhang "**Local Log Euclidean Multivariate Gaussian Descriptors and Its application to Image Classification**", IEEE Transaction on PAMI –Volume: 39, Issue: 4, pp.803-817, 2017.

9. Naik, B. Ramesh, and T. Venugopal. "**Object Recognition using Log Euclidean Multivariate Gaussian Descriptors**." International Journal of Applied Engineering Research, volume: 12, Issue:14 (2017): 4130-4137.

10. Zhang, Dengsheng, and Guojun Lu. "**Shape-based image retrieval using generic Fourier descriptor.**" Signal Processing: Image Communication 17.10 (2002): 825-848.

11. Zhao, Xiangyuan, and Brian Nutter. "**Content Based Image Retrieval system using Wavelet Transformation and multiple input multiple task Deep Autoencoder**." Image Analysis and Interpretation (SSIAI), 2016 IEEE Southwest Symposium on. IEEE, 2016.

12. Montazer, Gholam Ali, and DavarGiveki. "**Content based image retrieval system using clustered scale invariant feature transforms.**" Optik-International Journal for Light and Electron Optics 126.18 (2015): 1695-1699.

13. Grana, Costantino, et al. "**UNIMORE at ImageCLEF 2013: Scalable Concept Image Annotation**." CLEF (Working Notes). 2013.

14. Persoon, E. and Fu, K.S. '**Shape discrimination using Fourier descriptors',** IEEE Transactions on Systems, Man, and Cybernetics, Vol. 21, No. 3, March, pp.170–179. 1997

15. Hailong Shi, Hao Zhang, Gang Li,Xiqin Wang , " **Stable embedding of Grassmann Manifold via Gaussian Random Matrices**", IEEE Transaction on Information Theory, Vol. 61, No. 5 pp. 2924- 2924, May 2015.

16. MiquelCalvo., Josep M. Oller "**A distance between multivariate normal distributions based in an embedding into the Siegel group",** Journal of Multivariate Analysis Volume 35, Issue 2, pp. 223-242 November 1990.

17. V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "**Geometric means in a novel vector space structure on symmetric positive-definite atrices**," SIAM J. Matrix Anal. Appl, 2006.

18. J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek, "**Image classification with the Fisher vector: Theory and practice**," Int. J. Comput. Vis., vol. 105, no. 3, pp. 222–245, 2013.

19. D. L. Bihan, J. Mangin, C. Poupon, C. Clark, S. Pappata, and N. Molko, "**Diffusion tensor imaging: Concepts and applications,**" J Magn. Reson. Imaging, vol. 66, pp. 534–546, 2001.

20. D. Lowe, "**Distinctive image features from scale-invariant keypoints,**" Int. J. Comput. Vision, vol. 60, pp. 91–110, 2004.

21. G. Serra, C. Grana, M. Manfredi, and R. Cucchiara, "**Modeling local descriptors with multivariate Gaussians for object and scene recognition**," in Proc. ACMInt. Conf.Multimedia, 2013, pp. 709–712.

22. Y. Ke and R. Sukthankar, "**PCA-SIFT: a more distinctive representation for local image descriptors,**" in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2004, pp. II–506.

23. Ding, K., Xiao, L., &Weng, G. (2017). "**Active contours driven by region-scalable fitting and optimized Laplacian of Gaussian energy for image segmentation**". Signal Processing, 134, 224-233.

24. N. Dalal and B. Triggs, "**Histograms of oriented gradients for human detection,**" in Proc. Int. Conf. Comp. Vis. Patt. Recog., 2005, pp. 886–893.

25. Lu, S., Wang, S., & Zhang, Y. (2017). "**A note on the marker-based watershed method for X-ray image segmentation**". Computer methods and programs in biomedicine, 141, 1-2.

26. X. Pennec, P. Fillard, and N. Ayache, "**A Riemannian framework for tensor computing,**" Int. J. Comput. Vision, pp. 41–66, 2006.

27. Bai, X., Li, X., Fu, Z., Lv, X., & Zhang, L. (2017). "**A fuzzy clustering segmentation method based on neighborhood grayscale information for defining cucumber leaf spot disease images**". Computers and Electronics in Agriculture, 136, 157-165.

28. Bergstra, J. and Bengio, Y. "**Random search for hyper-parameter optimization**". Journal of Machine Learning Research, 13:281–305, 2012.

29. UIUC, Lecture 21. The Multivariate Normal Distribution, 21.5: **"Finding the Density"**.

30. Griffin, A. Holub, and P. Perona, "**The Caltech-256,"** California Institute of Technology, Tech. Rep., 2007.

31. Guy Lebanon, "**Metric Learning for Text Documents,**" IEEE transactions on pattern analysis and machine intelligence, vol. 28, no. 4, april 2006

32. Hall, B. C. **Lie groups, Lie algebras, and Representations: An Elementary Entroduction**. Graduate Texts in Mathematics. 222 (2nd ed.). Springer. doi:10.1007/978-3-319-13467-3. ISBN 978-3319134666 ISSN 0072-5285. *2015*

33. L. haran, R. Rosenholtz, and E. H. Adelson, "**Material perception: What can you see in a brief glance?**" J. Vis., vol. 9, no. 8, p. 784, 2009.

34. T. Kobayashi, "**Dirichlet-based histogram feature transform for image classification**," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2014, pp. 3278–3285.

35. S. Lazebnik, C. Schmid, and J. Ponce, "**Beyond bags of features: Spatial pyramid matching for recognizing**

natural scene categories," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2006, pp. 2169–2178.

36. Baker, **Matrix Groups: An Introduction to Lie Group Theory**. Springer-Verlag, 2002.
37. B. Hall, Lie Groups, **Lie Algebras, and Representations: An Elementary Introduction**. Springer, 2003.
38. L. T. Skovgaard, "**A Riemannian geometry of the multivariate normal model**," Scandinavian Journal of Statistics, vol. 11, no. 4, pp. 211–223, 1984.
39. J. Gallier, "**Logarithms and square roots of real matrices**," CoRR, vol. arXiv:0805.0245, 2013.
40. R. Gupta and A. Mittal, "**SMD: A locally stable monotonic change invariant feature descriptor,**" in Proc. Eur. Conf. Comput. Vis., 2008, pp. 265–277.
41. H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "**Speeded-up robust features (SURF),**"Comput. Vis. Image Underst., vol. 110, pp. 346–359, 2008.
42. Bhukya Ramesh Naik, VamsidharYendapalli, Naga Raju M, "Performance Improvement in CBIR using Region Weight Learning Approach", International Journal of Emerging Trends in Engineering Research, 8(7), July 2020, pp.3864-3869
43. Y. Huang, Z. Wu, L. Wang, and T. Tan, "**Feature coding in image classification: A comprehensive study,**" IEEE Trans**". Pattern Anal. Mach. Intell., vol. 36, no. 3, pp. 493–506, March 2014.
44. M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, and A. Vedaldi, "**Describing textures in the wild," in Proc. IEEE Conf. Comput. Vis.PatternRecognit**., 2014, pp. 3606–3613.
45. J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "**DeCAF: A deep convolutional activation feature forgeneric visual recognition**," in Proc. Int. Conf. Mach. Learn., 2014, pp. 647–655.
46. Ekeland, Ivar. "**The Hopf-Rinow theorem in infinite dimension."** Journal of Differential Geometry 13.2 (1978): 287-301.47]
47. K. van de Sande, T. Gevers, and C. Snoek, "**Evaluating color descriptors for object and scene recognition,**" IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 9, pp. 1582–1596, Sept 2010.