# MONOGENIC SCALE SPACE BASED REGION COVARIANCE MATRIX DESCRIPTOR FOR FACE RECOGNITION

M. Sharmila Kumari[*], B H Shekar[#] and G Thippeswamy[$]

[*]Department of Computer Science and Engineering, P A College of
Engineering, Mangalore, Karnataka, India.
[#]Department of Computer Science, Mangalore University, Karnataka, India.
[$]Department of Computer Science and Engineering, Sir MVIT, Yalahanka,
Bangalore, Karnataka, India.
`{sharmilabp, bhshekar, swamy.gangappa}@gmail.com`

## ABSTRACT

*In this paper, we have presented a new face recognition algorithm based on region covariance matrix (RCM) descriptor computed in monogenic scale space. In the proposed model, energy information obtained using monogenic filter is used to represent a pixel at different scales to form region covariance matrix descriptor for each face image during training phase. An eigenvalue based distance measure is used to compute the similarity between face images. Extensive experimentation on AT&T and YALE face database has been conducted to reveal the performance of the monogenic scale space based region covariance matrix method and comparative analysis is made with the basic RCM method and Gabor based region covariance matrix method to exhibit the superiority of the proposed technique.*

## KEYWORDS

*log-Gabor Transform, Riesz transform, Monogenic scale space, Region covariance matrix descriptor, Face recognition.*

## 1. INTRODUCTION

The biometrics is the well addressed research field in the domain of computer vision and we have seen plethora of algorithms in particular on face recognition because of its wide acceptability in several applications ranging from identity authentication, access control and face based video indexing/retrieval to human–computer interaction/communication. In the recent days, among the several models proposed for face recognition, descriptor based algorithms gain much importance because of their robustness against illumination, noise, and occlusion – the common problems encountered in real recognition environment.

The local descriptors [8, 14, 22] are commonly employed in a number of real-world applications such as object recognition [5, 14] and image retrieval [16] as they can be computed efficiently, are resistant to partial occlusion, and are relatively insensitive to changes in viewpoint. Mikolajczyk and Schmid [15] presented a comparative study of several local descriptors including steerable filters [7], differential invariants [10], moment invariants [22], Scale Invariant

Feature Transform [13], and cross-correlation of different types of interest points [8, 16]. Their experiments showed that the ranking of accuracy for the different algorithms was relatively insensitive to the method employed to find interest points in the image but was dependent on the representation used to model the image patch around the interest point. In the SIFT algorithm, each keypoint is represented by its neighbourhood, described as a set of orientation histograms computed from the gradient image. The SIFT descriptors are invariant to scale, rotation, lighting and viewpoint change. The most common implementation uses a 128 dimensional descriptor. Ke and Sukthankar [9] proposed PCA-SIFT descriptor which is also based on the gradient image, the main difference with SIFT being the further compression using PCA. Recently, Speeded Up Robust Features (SURF) descriptor [1] has appeared as an alternative to SIFT. Its main advantage is its fastest computation, while keeping a high descriptive power. It is partially inspired by SIFT, but instead of using the gradient image, it computes first order Haar wavelet responses. The LESH or Local Energy based Shape Histogram has been specifically designed for face recognition applications. Its goal is to encode the underlying shape present in the image. Basically, the descriptor is a concatenation of histograms obtained by accumulating local energy along several filter orientations.

On the other hand, we have seen appearance based approaches which are capable of withstanding noise and occlusion and simple in terms of implementation. In these approaches, data transformation is a fundamental step and the goal is to obtain highly discriminative lower-dimensional data from high-dimensional data. Principal component analysis (PCA) and linear discriminant analysis (LDA) are the widely used tools in the face recognition domain, which encode high-dimensional face images as lower-dimensional *eigenfaces* [20] and *fisherfaces* [2] respectively. PCA is a linear method that ensures that the data transformed are uncorrelated and preserve maximally the second order statistics of the original data, and hence is insensitive to the dependencies of multiple features in the patterns. To overcome this problem, kernel PCA [19] is proposed as a non-linear extension of PCA that computes the principal components in a high-dimensional feature space. On the similar line, kernel FLD is proposed as a non-linear extension to linear discriminant analysis.

We have also seen the frequency domain based approaches where high-frequency components are used as facial features because of their robustness to illumination changes. The local phase information is used in some techniques as it is proved to be sufficient to completely reconstruct a signal within a scale factor [17]. As the Gabor wavelet representation captures salient visual properties such as spatial localization, orientation selectivity, Lades *et al.* [11] applied Gabor wavelets for face recognition using dynamic link architecture framework. Wiskott *et al.* [23] extended this basic framework to devise Gabor wavelet based elastic bunch graph matching method to label and recognize faces. Liu [12] proposed Gabor-based kernel PCA method by integrating the Gabor wavelet representation of face images and the kernel PCA method for face recognition. Porikli and Tuzel [18] and Tuzel *et al.* [21] proposed a new descriptor framework called region covariance matrices (RCM) for object detection and tracking. The RCMs can be categorized as a matrix-form feature. However, direct application of RCM to face recognition has not produced satisfactory results and hence Yanwei *et al.* [24] introduced Gabor-based region covariance matrices as face descriptors. The results of Yanwei *et al.* [24] are highly encouraging when compared to any other Gabor wavelet based techniques. Although Gabor wavelet based face recognition models possess very high recognition accuracy irrespective of noise, occlusion and illumination problems, massive computing and space requirements are the major bottleneck in these approaches. In this context, we have proposed a new frame work for face recognition based on monogenic scale space model. The monogenic signal requires fewer convolutions when compared to Gabor wavelet based model and hence require much lesser computing time. The region covariance matrix descriptor is employed on monogenic scale space based images to obtain the representation of face images.

The remaining part of the paper is organized as follows. In section 2, we review monogenic scale space technique. The region covariance matrix descriptor and its construction in monogenic scale space for face recognition is presented in sections 3 and 4. The experimental results and comparative study is given in section 5. Conclusion is reached in section 6.

## 2. MONOGENIC SCALE SPACE: A REVIEW

The monogenic signal [3, 4] is based on the Riesz transform which is used instead of the Hilbert transform. The monogenic signal analysis [3, 4] is a framework to interpret images in terms of the local phase, local orientation and local energy. The monogenic signal is an effective tool to analyze 2-D signals in a rotation invariant manner. The signal is built upon the first order Riesz transform. The spatial representation of the Riesz kernel in 2D space is:

$$\left(R_x(\mathbf{x}), R_y(\mathbf{x})\right) = \left(\frac{x}{2\pi|\mathbf{x}|^3}, \frac{y}{2\pi|\mathbf{x}|^3}\right), \mathbf{x} = (x,y)\epsilon\mathbb{R}^2 \tag{1}$$

and its transfer function in the Fourier domain is:

$$(F_u(\mathbf{u}), F_v(\mathbf{u})) = \left(-i\frac{u}{|\mathbf{u}|}, -i\frac{v}{|\mathbf{u}|}\right), \mathbf{u} = (u,v)\epsilon\mathbb{R}^2 \tag{2}$$

For any image, say I($\mathbf{x}$), the monogenic signal is defined as the combination of I and its Riesz transform:

$$\text{i.e., } I_m(\mathbf{x}) = (I(\mathbf{x}), R_x\{I\}(\mathbf{x}), R_y\{I\}(\mathbf{x}))$$
$$= (I, R_x * I, R_y * I) \tag{3}$$

where * stands for the convolution operation.

And hence the local orientation is calculated as [3, 4]:

$$\theta = \arctan\frac{R_y\{\mathbf{I}\}}{R_x\{\mathbf{I}\}}, \theta \epsilon [0,\pi) \tag{4}$$

The local phase is defined as:

$$\varphi = atan2\left(\sqrt{R_x^2\{\mathbf{I}\} + R_y^2\{\mathbf{I}\}}, \mathbf{I}\right), \varphi\epsilon[0,\pi) \tag{5}$$

The local energy is defined as:

$$E = \left(\sqrt{R^2\{\mathbf{I}\} + R_x^2\{\mathbf{I}\} + R_y^2\{\mathbf{I}\}}, \mathbf{I}\right) \tag{6}$$

where $R^2\{I\} = I * F^{-1}(G(w))$. Here G(w) is the log-Gabor filter in the Fourier domain.

Since log-Gabor filters are band-pass filters, usually multi-scale monogenic representation is required to fully describe a signal. In figure 1, we have given the convolved images in monogenic scale space filter showing the log-Gabor transformed image with its energy and orientation images. One can see that the local structure is well captured in monogenic components.
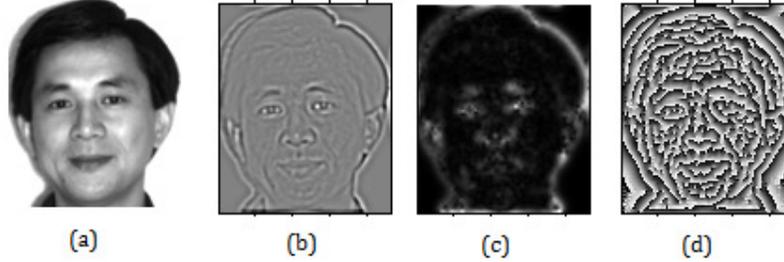


**Fig. 1.** (a) Original image; (b) log-Gabor image; (c) Energy image and (d) Orientation image

## 3. REGION COVARIANCE MATRIX: A REGION STATISTICS BASED DESCRIPTOR

The region covariance matrix (RCM) proposed by Tuzel *et al.* [21] is an efficient matrix form descriptor which is defined as the covariance of many image statistics computed inside the region of an image. Let I be an image of size MxN. Define a function $\phi$ that extracts $d$ dimensional feature vector $z_i$ from a pixel at (x, y) of I. i.e.,

$$\phi(I, x, y) = z_i \in R^d \tag{7}$$

where $i = y \times M + x$ is the index of (x, y). Considering all the pixels in a region R, it is represented as dxd covariance matrix ($C_R$) of the feature points $z_i$.

$$i.e., C_R = \frac{1}{n-1} \sum_{i=1}^{n} (z_i - \mu_R)(z_i - \mu_R)^T \tag{8}$$

where $\mu_R$ is the mean of $z_i$

In [21], the function $\phi$ is defined by the pixel location, intensity value, (colour values (RGB) in case of colour image) and the norm of the first and second order derivatives of the image.

$$i.e., \phi(I, x, y) = [x \quad y \quad I(x,y) \quad |I_x| \quad |I_y| \quad |I_{xx}| \quad |I_{yy}|], \tag{9}$$

where $|.|$ denotes the absolute operator, $I_x$ and $I_y$ respectively denote the first order partial derivative wrt x and y and, $I_{xx}$ and $I_{yy}$ denote the second order partial derivative wrt x and y. One can notice here that the RCM is a symmetric matrix and capable of capturing both spatial and statistical properties as diagonal entries represent the variance of each feature and non-diagonal entries represent their respective correlations.

For classification purpose, the distance measure proposed by Forstner and Moonen [6] is used which is an eigen-value based distance measure.

$$\partial(CR_1, CR_2) = \sqrt{\sum_{i=1}^{d} \ln^2\big(\lambda_i(CR_1, CR_2)\big)} \qquad (10)$$

where $\lambda_1, \lambda_2, \dots \lambda_d$ are the generalized eigenvalues of $CR_1$ and $CR_2$ computed from:

$$\lambda_i CR_1 u_i = CR_2 u_i, \; i=1,2,\dots,d.$$

Although RCM based methods are successful in people detection, object tracking, texture classification etc., they exhibit poor performance for face recognition problems. Hence, we proposed to extract monogenic signal based RCM, which possess better recognition rate when compared to the state of the art face recognition algorithms.

## 4. CONSTRUCTION OF REGION COVARIANCE MATRIX DESCRIPTOR IN MONOGENIC SCALE SPACE.

The monogenic scale space based features are obtained by employing Riesz transform. The monogenic filter is employed on an intensity image, say **I** at different scales to obtain the local energy, local phase and local orientation information. In our work, we have considered local energy at five different scales for constructing RCM which is as follows.

It shall be observed from Eq. (6) that the energy can be computed at different scales. Let $A_1,\dots,$ $A_5$ be the energy computed at five different scales. Hence, the feature mapping for each pixel (x, y) is defined as:

$$\phi(I, x, y) = [A_1(x, y)\; A_2(x, y)\; A_3(x, y)\; A_4(x, y)\; A_5(x, y)]$$

It shall be observed here that the size of the RCM is 5x5 = 25. One can notice here that the proposed model requires only five convolutions with the Riesz transform where as Gabor wavelet based representation requires 40 convolutions (5 orientation and 8 scales) with Gabor filter and hence the feature mapping results in 40x40 dimension feature vector which is very massive and hence consume much memory to store the feature vector. Obviously the computing time increases due to more number of faces in image databases.

So far, we have considered whole image as an input. It is possible to take upper half, lower half, left half or right half of an image and could construct RCM possibly to handle occlusion.
As suggested in [19], the eigen-value based distance measure (Eq. (10)) is considered for finding the similarity between two faces images.

## 5. EXPERIMENTAL RESULTS

This section presents the results of the experiments conducted to corroborate the success of the proposed model. We have conducted experimentation using two benchmark face image databases namely AT&T and YALE face databases. All the experiments are conducted using P-IV machine on Windows-7 platform using MATLAB 7.8 tool.

**Experimentation on AT&T face databases:** The AT&T face database consisting of 400 gray-scale images of 10 subjects, each covering a wide range of poses as well as race, gender and

appearance. In Fig. 2, we have shown subset of one such subject of the AT&T database. The experimentation consists of varying number of training samples and testing samples under each class. We have chosen first five samples for training and the remaining samples for testing for each person and the recognition accuracy is obtained. Similarly, two other test cases are generated which consists of first four and three samples for training and the remaining samples for testing. The results obtained due to the proposed model are reported in Fig. 3. The results obtained due to Gabor wavelet based RCM and basic RCM are also shown in Fig. 3. It shall be observed that the recognition accuracy of the proposed model is on par with the Gabor wavelet based RCM, and quite high when compared to the basic RCM approach. Over all, the performance of the proposed model is good as it consumes less time with good recognition results.



**Fig. 2.** Subset of one subject of AT&T face database
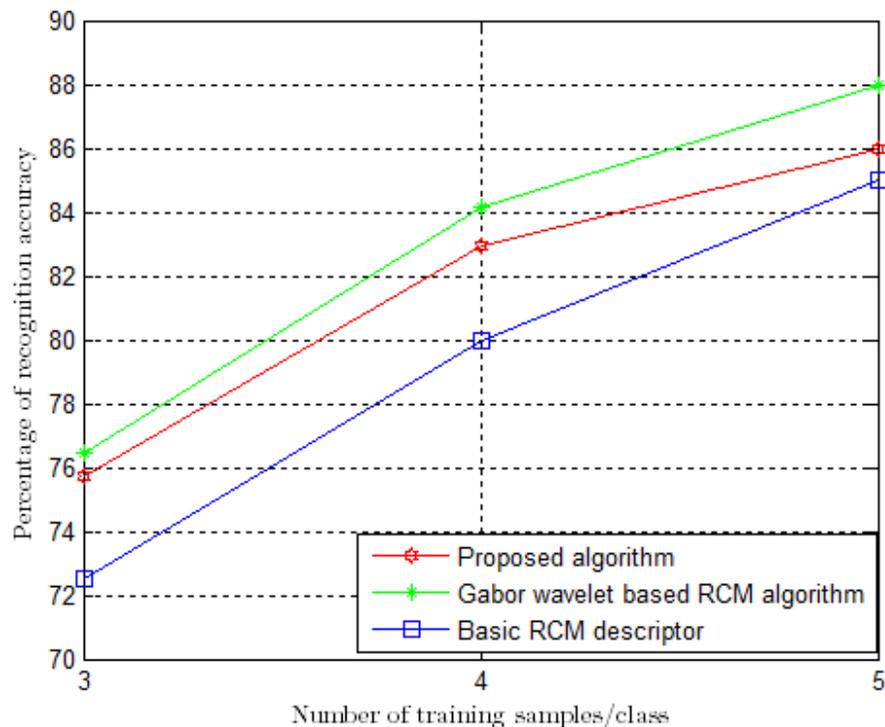


**Fig. 3.** Recognition accuracy of the proposed model and other techniques on AT&T face database.

**Experimentation on YALE face databases:** The YALE face database contains 165 images of 15 subjects that include variation in both facial expression and lighting. In Fig. 4, we have shown the closely cropped images which include internal facial structures. Here, we have made the following type of testing. The training set comprised of six images randomly chosen for each person with remaining number of face images for each person. The recognition accuracy is computed for the proposed model, Gabor wavelet based RCM technique and the basic RCM

technique. Similarly, four and five face images are chosen randomly under each person during training and the remaining face images are used for testing. The results are shown in Fig. 5. It shall be observed from Fig. 5 that the proposed model posses better performance.



**Fig. 4.** Subset of one subject of YALE face database
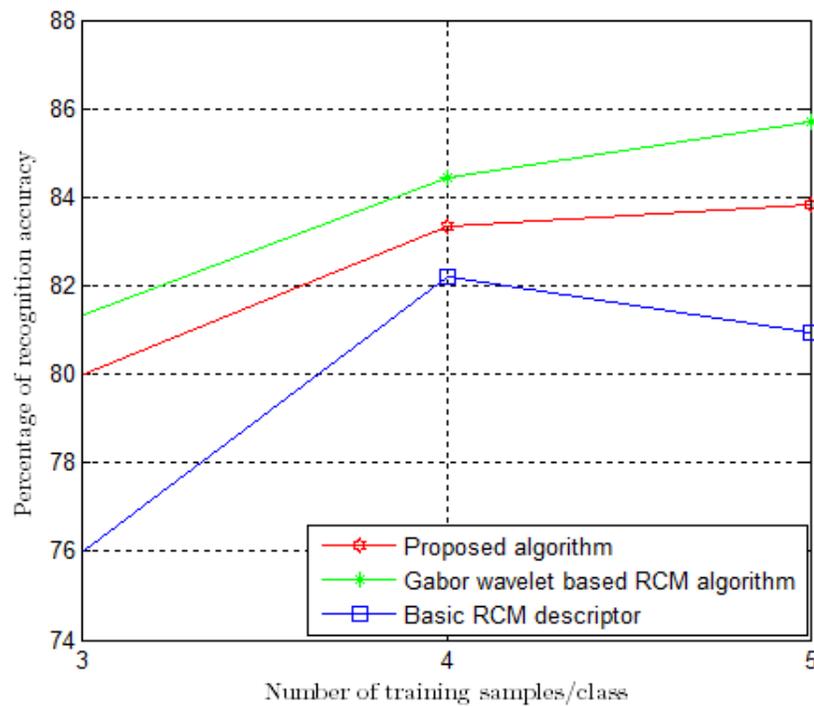


**Fig. 5.** Recognition accuracy of the proposed model and other techniques on YALE face database.

## 6. CONCLUSION

The monogenic scale space based region covariance matrix is developed for face recognition. The energy information contained in the monogenic signal at different scales is used to build the region covariance matrix descriptor for each face image. Experimental results on the standard benchmark databases reveal the superiority of the proposed model for face recognition problems and its suitability in real environment for face recognition problems.

# REFERENCES

1. Bay H, Ess A, Tuytelaars T, Gool L V, (2008). "Speeded-Up Robust Features (SURF)", *Computer Vision and Image Understanding*, Vol. 110, Issue 3, pp. 346-359.
2. Belhumeur, P.N., J.P. Hespanha, and D.J. Kreigman, (1997). "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19(7), pp. 711-720.
3. Felsberg, M. and Sommer, G. (2004). "The monogenic scale-space: A unifying approach to phase-based image processing in scalespace", *Journal of Mathematical Imaging and Vision*, 21(1), pp.5–26.
4. Felsberg, M., and G. Sommer, (2001). "The monogenic signal", *IEEE Transactions on Signal Processing*, vol. 49, pp. 3136-3144.
5. Fergus, R., P. Perona, and A. Zisserman, (2003). "Object class recognition by unsupervised scale-invariant learning", *In Proceedings of Computer Vision and Pattern Recognition*, pp. 264-271.
6. Forstner, W. and Moonen, B, (1999). "A metric for covariance matrices", *Technical report*, Dept. of Geodesy and Geoinformatics, Stuttgart University.
7. Freeman, W.T. and E. H. Adelson, (1991). "The design and use of steerable filters", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13(9), pp. 891–906.
8. Harris, C. and M. Stephens, (1988). "A combined corner and edge detector", *In Alvey Vision Conference*, pages 147–151.
9. Ke, Y. and R. Sukthankar, (2004). "PCA-SIFT: A more distinctive representation for local image descriptors", *In Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, Washington, pp. 506-513.
10. Koenderink, J. and A. van Doorn, (1987). "Representation of local geometry in the visual system", *In Biological Cybernetics*, Vol. 55, pp. 367–375.
11. Lades, M., J.C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Wurtz, and W. Konen, (1993). "Distortion Invariant Object Recognition in the Dynamic Link Architecture", *IEEE Transactions on Computers*, vol. 42, pp. 300-311.
12. Liu, C., (2004). "Gabor-Based Kernel PCA with Fractional Power Polynomial Models for Face Recognition", *IEEE Transactions on Pattern Analysis And Machine Intelligence,* Vol. 26 (5), pp. 572-581.
13. Lowe, D.G., (2004). "Distinctive image features from scale-invariant keypoints", *International Journal of Computer Vision*, vol. 60(2), pp. 91-110.
14. Lowe, D.G., (1999). "Object recognition from local scale-invariant features", *In Proceedings of International Conference on Computer Vision*, pp. 1150–1157.
15. Mikolajczyk, K and C. Schmid, (2005). "A performance evaluation of local descriptors", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27(10), pp. 1615-1630.
16. Mikolajczyk, K and C. Schmid, (2001). "Indexing based on scale invariant interest points". *In Proceedings of International Conference on Computer Vision*, pages 525–531.
17. Oppenheim, A.V. and 1. S. Lim, (1981). "The importance of phase in signals", *Proceedings of The IEEE*, vol. 69, no. 5.
18. Poriklim F and O. Tuzel, (2006). "Fast construction of covariance matrices for arbitrary size image windows", *In Proceedings of IEEE International Conference on Image Processing*, pp. 1581-1584.
19. Schokopf, B., S. Mika, C.J.C Burges, P. Knirsch, K.-R Muller, G. Ratsch, and A.J. Mola, (1999). "Input space versus feature space in kernel based methods", *IEEE Transactions on Neural Networks*, vol. 10(5), pp. 1299-1319.
20. Turk, M. and A. Pentland, (1991). "Eigenfaces for recognition", *Journal of Cognitive Neuroscience*, vol. 3(1), pp. 71-86.
21. Tuzel, O., F. Porikli, and P. Meer, (2006). "Region covariance: a fast descriptor for detection and classification", *In the Proceedings of European Conference on Computer Vision*, pp. 589-600.
22. Van Gool, L., T. Moons, and D. Ungureanu, (1996). "Affine/photometric invariants for planar intensity patterns", *In Proceedings of European Conference on Computer Vision.* pp. 642-651.
23. Wiskott, L., J.M. Fellous, N. Kruger, and C. von der Malsburg, (1997). "Face Recognition by Elastic Bunch Graph Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775-779.
24. Yanwei P., Yuan Yuan and Xuelong Li, (2008). "Gabor-based region covariance matrices for face recognition", *IEEE Transactions on circuits and systems for video technology*, vol. 18(7), pp. 989-993.

**AUTHORS**

**M. Sharmila Kumari** obtained her B.E., degree from University of Mysore in the year 2000 and M.Tech., degree in Computer Science and Technology from Visweswaraya Technological  University in the year 2004. She is currently working as an Associate Professor in the Department of Computer Science and Engineering, PA College of Engineering, Mangalore, India. She has authored about twenty five peer-reviewed papers in International Journals and Conferences. Her areas of research cover Video Image Processing, Biometrics and Content Based Image Retrieval.

 **B. H. Shekar** obtained his B.Sc., M.Sc., and Ph.D., degrees in Computer Science and Technology from the University of Mysore, Mysore, India, respectively in the years 1992, 1994, and 2007. He is currently working as an Associate Professor in the Department of Studies in Computer Science, Mangalore University, Mangalore. He has authored about 15 Journal papers and 50 peer-reviewed International Conference papers. His areas of research cover Video Image Processing, Biometrics, Object recognition and Shape analysis.

**G. Thippeswamy** obtained his B.E., and M.E., degrees in Computer Science and Technology from Visweswaraya Technological University. He is currently working as an Assistant Professor in the Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology, Yalahanka, Bangalore, India. He has authored about fifteen peer-reviewed papers in International Journals and Conferences. His areas of research cover Biometrics and Content Based Image Retrieval.