Energy Interpolated Template Coding for Video Compression in Traffic Surveillance Application

Shivprasad P. Patil[1], Rajarshi Sanyal[2] and Ramjee Prasad[1]

[1]*Department of Business Development and Technology, Aarhus University, Herning, Denmark*
[2]*Belgacom International Carrier Services, Brussels, Belgium*
E-mail: sppatil1212@gmail.com; ramjee@btech.au.dk; rajarshi.sanyal@bics.com

Received 26 January 2018; Accepted 27 April 2018;
Publication 3 July 2018

Abstract

In video coding, exploitation of temporal correlation between frames is an important step for reduction of redundant data in successive video frames. However, dynamic nature of video content introduces difficulty in finding temporal correlation. In this paper we propose a novel template coding approach to compress the video data for traffic surveillance which addresses above said difficulty. In this work, the conventional approach of the template coding, wherein two successive frames are considered, is improved by a 'dynamic model' of the template. The dynamism of template selection is achieved through energy interpolation of successive frame data over some time period, rather than only two successive frame data. A coherent histogram model is developed to build accurate template to achieve improvement in compression. The proposed efficient template matching approach predicts exact template thereby minimizing the processing overheads and reduction in processing time. The obtained simulation result unveils that, the proposed approach results in accurate template localization, thereby improving the accuracy in coding and the coding speed in comparison to conventional template based compression approaches.

Journal of Mobile Multimedia, Vol. 14_3, 257–272.
doi: 10.13052/jmm1550-4646.1431

Keywords: Video compression, Template coding, Energy interpolation, Traffic surveillance.

1 Introduction

Automation in real time applications has gained a lot of interest in recent past. Video coding for information retrieval, monitoring, controlling is one such application and which has been investigated in multiple direction for its automation. With the emerging issue of security concerns and human negligence, various issues are surfacing, which could have been stopped. The issue is more prominent as various application demands for continuous monitoring to take immediate action. In most surveillance applications, capturing devices such as digital cameras are installed at a remote monitoring zone and the captured information's are streamed to monitoring location via dedicated cable link or via wireless link. Where cable links are more effective in accurate communication and transmission rate, the range constraint and their installation and maintenance cost are always a limitation. In wireless mode communication, interference, resource scarcity such as transmitting bandwidth is the main limiting constraint. To overcome this issue of video transmission with resource scare wireless system, various approaches of compression model were developed in past. An annotation based coding for video information prediction is presented in [1]. The system uses a motion parameter for prediction of template base to derive matching template. But it is very difficult to accurately model the base template due to large diversity of video data. A hidden markov approach to template based coding is outlined in [2]. A global and local temporal modeling is used as a motion detail for template representation. However, recognition rate depends on training data and test data. In [3] motion energy based coding is developed. A motion based coding using energy and the variation in frame content (history) is used for the compression. In this approach, motion component due to camera motion are not separated out. A local representation of the extracted spatio-temporal interest points (STIPs) [4, 5] from a video sample is used for video compression. However, this work does not support for camera motion. Also, there is no mention on processing speed. The local representation is observed to be robust to statistical representation of video details outlined in [6, 7]. In motion detection model, Harris detector [8] and 3-D SIFT [9, 10], are used for the localization of non moving parameter defined as matching template. The SIFT operator performs a max/min searching over the difference of Gaussian (DoG) function for each video frame. However, in the scenario of cluttered

video, these approaches are not suitable. Though there is no mention on time complexity, it seems that they are computationally costly.

A combined format of histogram oriented gradient and histogram optical flow (HoG-HoF) for motion detection and compression is outlined in [11]. The histogram of gradient is observed to be an effective approach for motion detection model. A 3D- HoG [12] is outlined, with a set of descriptive approach to obtain effective motion model. Towards optimization of motion detection model, a histogram based coding [13] for video retrieval is developed. The system uses the localized temporal histogram templates to detect a motion model from a video dataset. However in these cases, camera motion is not taken in to the consideration, thereby introducing less accuracy in motion vector detection. Further, in most of the existing approach using energy interpolation, the definition of templates is based on histogram details. However, all the past developments are confined to a predefined template. Wherein no efforts were made in defining the matching template based on the frame content and frame variations. In this paper, a new approach to template based video compression is proposed, where energy interpolation is used for frame correlation to derive energy difference and the correlative factor is then used as a defining parameter for template development. The rest of the paper is organized as follows. Section 2 outlines the conventional template based coding for video compression. Section 3 outlines the proposed energy interpolation model for dynamic template definition. The obtained simulation results are presented in Section 4. Section 5 concludes the presented work.

2 Template Based Coding

In various approaches of video compression, template based coding is preferred for video compression due to its lower computational complexity and thereby fast processing. Among different approach, an approximate nearest neighbor (ANN) was presented in [14]. This approach present a data portioning based on k-mean clustering and hierarchical extension to derive a template block. The method process on overlapped block size and derive a residual error based on the direct comparison of the two successive frame information's. In [15] a template match predication (TMP) was defined to develop compression model. The TMP method derive the self similarity content of the two successive frame details and derive the frame correlation based on the distance measure of self and neighbor pixels. This model uses the static pixel details to derive a template match. In this approach no motion details are used. To improve the template mapping in [13], an energy interpolation of static frame

content using histogram is developed. This method defines a temporal and spatial localization of template based on histogram mapping. In the process of temporal coding, a set of histogram defining the temporal template for different slices of the frames is developed. The developed histogram correlation (E_d) is categorized into a set of temporal matching histogram (E_c), where $E_c \, \epsilon \, E_d$. The categorization of templates hence leads to faster localization of template model in the search process of motion element in a given video sample (E_s). In the search process an intersection of histogram template for temporal localization is used. In such coding, the intersection of the histogram is defined by;

$$I\left(E_s, E_d\right) = \sum_{i=1}^{m} \left(\frac{\min\left(E_s^i - E_d^i\right)}{E_N^i} \right) \tag{1}$$

Where, E_N is the normalized histogram for the observing video. A major assumption in this model is that a high value for $I\left(E_s, E_d\right)$ is predictive that time frame d contains part of frame s.

Here it is to be observed that template categorization is carried out using a temporal template defined in dataset, and the dynamicity of the video content is neglected. Here, dynamicity is referred to nonlinear variations in video contents due to unpredictable vehicle flow. It is worth to note that maximum energy contents are present in these dynamic contents and if it is not considered, it may lead to wrong template definition. Further, in various video samples the variations are non redundant, such as the foreground static-background moving, or foreground and background moving, background static and foreground moving, and various application where cameras are moved, wherein a false motion is observed (due to motion of the camera). In such scenario, the template differentiation needs to be dynamic to achieve higher accuracy and faster processing.

3 Energy Interpolated Template Coding

In conventional energy based template approach, two successive frames are considered for processing purpose. It is evident that, the accuracy of this approach will not be adequate if video contents are changing dynamically. Therefore, it is essential to have a dynamic template model to overcome the issues arising from distortions in video data. In this work a dynamic template model based on the video content is derived through few successive frames

histogram correlation. A multi frame inter correlative frame error is computed based on the recurrent frame histogram correlation. In this approach, for a given video sample each of the frame energy is computed and an energy set is defined as,

$$E_i\left(m\right) = \left[E\left(mN\right),\ E_i\left(mN-1\right), \ldots.., E_i\left(mN\right)\right] \tag{2}$$

Where, E_i is the energy histogram for each frame of a video file. N is number of frame. To compute the frame error (F) for the given two frame data, where the energy interpolates are compared to derive energy difference is defined as;

$$F_{i,E}\left(m\right) = E_{i,t}\left(m\right) - E_{i,t+1}\left(m\right) \tag{3}$$

The set of frame correlation on energy plane is computed, and from the obtained frame errors, an optimal value is selected for the satisfying condition of $\min(F_{i,E}\left(m\right))$.

To develop the frame interpolation template, a reference energy histogram is derived from a successive frame data observed over a period of time. To optimize the template computation, a histogram normalization process is made by a weight parameter defined by,

$$E_i\left(k\right) = E_i\left(k\right)v(k) \tag{4}$$

Where, $v\left(k\right) = \left[v_0\left(m\right),\ v\left(m\right), \ldots\ldots, v_{M-1}\left(m\right)\right]^T$ are set of frame weight defined for each frame. The values are randomly initialized and propagated to a minimal error value in a recurrent manner. The Frame error is then modified as,

$$F_{i,E}\left(m\right) = E_{i,t}\left(m\right) - E_i\left(m\right)v(m) \tag{5}$$

An initial error value of frame error $F_{i,E,0}$ is recorded and the frame error is updated for all successive frames. At each of the iteration the weight values are updated based on the histogram energy and the frame error, as given below

$$v\left(m+1\right) = v\left(m\right) + \mu\sum\nolimits_{i=0}^{N-1} \frac{E_i^T(m)}{\left\|E_i(m)\right\|^2} F_{i,E,0}\left(m\right) \tag{6}$$

The weight value is updated with a step size μ which is used to govern the weight value to a constant updation rather than a random updation.

The deviation in the bin variation of the histogram is then integrated over a period of 0 to N defined by,

$$\mathrm{D}(E_{i,N}) = \int_{\mathbf{0}}^{\mathbf{N}} \mu \sum_{i=0}^{N-1} \left(2E \left[\frac{E_{i,n}(m)\tilde{v}\,(m)\,F_{i,N}\,(m)}{\|E_{i,N}(m)\|^2} \right] - \mu E \left[\frac{e_{i,NE}^2(m)}{\|E_{i,N}(m)\|^2} \right] \right) \tag{7}$$

Wherein integrating the estimate, over 'N' observation period accumulates the estimation of 'N' inter frame errors. For each frame with minimum estimate error is then selected as the selected histogram bin and an intersection bin is then derived from Equation (1). This template selection result in the selection of template region for any motion parameter with minimum correlative frame error. The suggested approach is interpreted in Figure 1.

A k-nearest neighbour model is used for the interpolation of the frame back in decoder unit. For each of the template region a signature is used with motion parameter, and the marked template region is interpolated by the replication of reference template to regenerate the video frame. To validate the proposed approach, a simulation is carried out for the proposed system in comparison to conventional energy model and template based compression model.

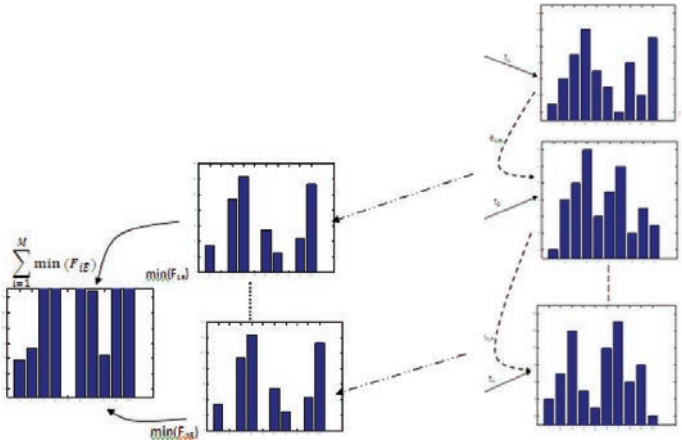

Figure 1 The energy correlative template selection approach.

4 Simulation Results

The proposed system is developed in Matlab 8.1 and tested for a real time traffic surveillance application. The day time test sample is captured from city of Pune (India) traffic flow using a rotating capturing camera installed over a junction point. The video samples are captured at 25fps. A frame sample of captured video is shown in Figure 2.

For the processing of the captured video sample, a frame reading is carried out at a frame time skip of ten frames to obtain dominant motion details. The extracted frames for processing are shown in Figure 3.

The frames are selected to have three distinct motion parameter, where the foreground is observed to be moving consisting of moving vehicles, the background are static reflecting the sign boards, tree and a false motion model due to camera motion is taken which gives a motion effect to static roads, footpath and static surroundings. A template extraction to this captured frame is carried out using Template match prediction (TMP) [15] and Histogram energy based template matching (HIST) [13] compared with the proposed energy interpolated template matching (EI-HIST).

Figure 2 Captured sample of a traffic surveillance camera.

Figure 3 Extracted frames for processing.

For the TMP template matching, the template used for compression model is derived based on the comparison of successive frames. The obtained template for mapping is shown in Figure 4.

For the prediction of template more effectively, energy prediction model using histogram feature as outlined in [13] is developed. The obtained template coefficient using Histogram energy model (HIST) is shown in Figure 5.

The obtained template for video compression using the proposed energy interpolated HIST (EI-HIST) is shown in Figure 6. The Finest coarseness in the selection of moving element based on the interpolation approach could be observed. The finer prediction results in lower coefficients, thereby achieving the compression.

The processing accuracy and computation overhead is computed in terms of accuracy in retrieved frame data and the time taken for the encoding and decoding of the developed system. The derived frame by the interpolation using TMP, HIST and EI-HIST is shown in Figures 7, 8 and 9 respectively.

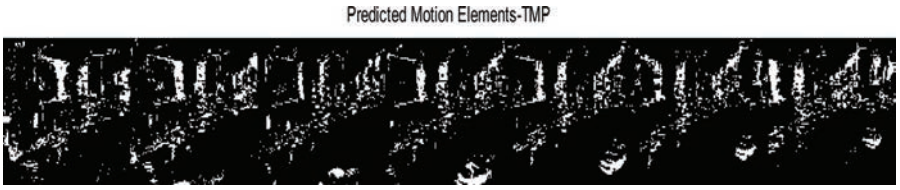

Figure 4 TMP [15] based template coefficient.

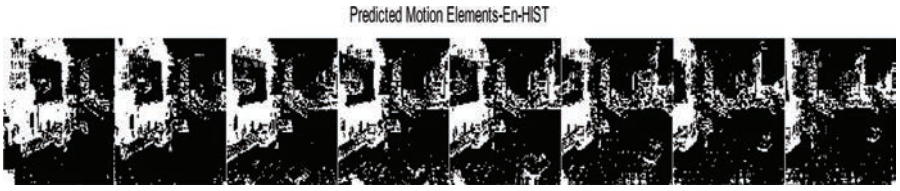

Figure 5 Template derived using Histogram mapping [13].

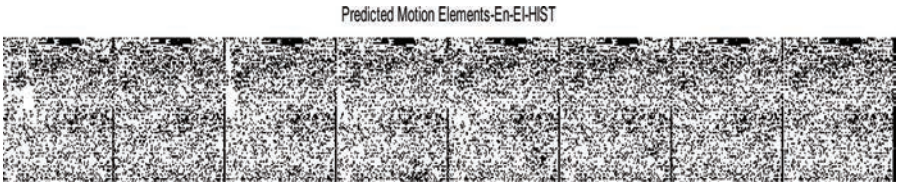

Figure 6 Template derived by EI-HIST.

The performance of the developed system is measured w.r.t. the actual motion element detected. The coefficient count for the conventional method and proposed method is outlined in Table 1.

The accuracy of these three developed methods are compared by the obtained PSNR,

$$PSNR = 10\log\left(\frac{[\max(I)]^2}{MSE}\right) \tag{8}$$

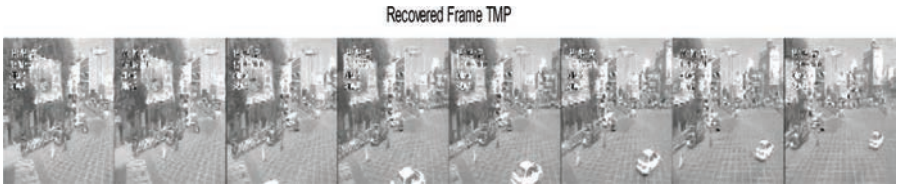

Figure 7 Recovered frame using TMP approach.

Figure 8 Recovered frame using HIST approach.

Figure 9 Recovered frame using EI-HIST approach.

Table 1 Motion coefficients accounted by the three developed methods

Method / Parameter Value	TMP [15]	HIST [13]	EI-HST
Original Sample Size	765952	765952	765952
Redundant coefficients	597832	560750	223068
Motion Element detected	168120	205202	542884
Data Overhead	56.94%	46.65%	17.63%

Where,

$$MSE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (I - I')^2 \tag{9}$$

Where I and I' are the original and interpolated frames.

The comparison of the PSNR obtained for the developed method is shown in Figure 10. The PSNR for the energy interpolation domain is observed to be high due to spectral domain processing, rather than time domain processing. In case of TMP approach, coding part accumulates more redundant data, which leads to error in decoding process, resulting in lower PSNR value.

The template approach affects the computation time based on the derived template region. The accurate region detection result in higher interpolation accuracy and low computation time. The measured time delay for processing, processed over a Intel(R) core i5 CPU at 2.3GHz processor, is shown in Figure 11. An improvement of 0.3 Sec and 0.8 sec is observed in comparison to the conventional Histogram interpolation and TMP approach respectively.

The observed overhead in processing data for the three methods is presented in Figure 12. The elimination of data redundancy results in minimization of data overhead in the computation. It is observed to be reduced by 39% and 29% in comparison to TMP and HIST approaches respectively.

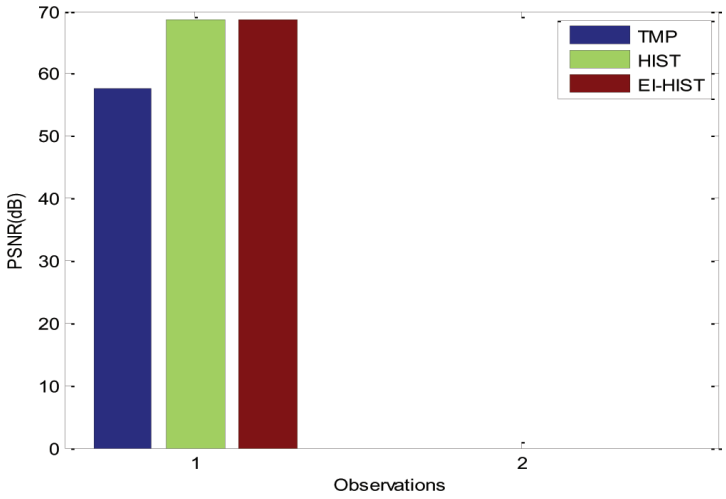

Figure 10 PSNR comparison for the developed approach.

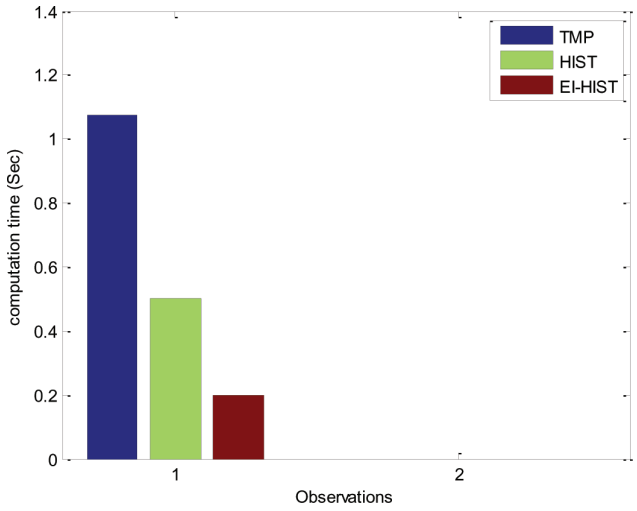

Figure 11 Computation time plot for the three developed approach.

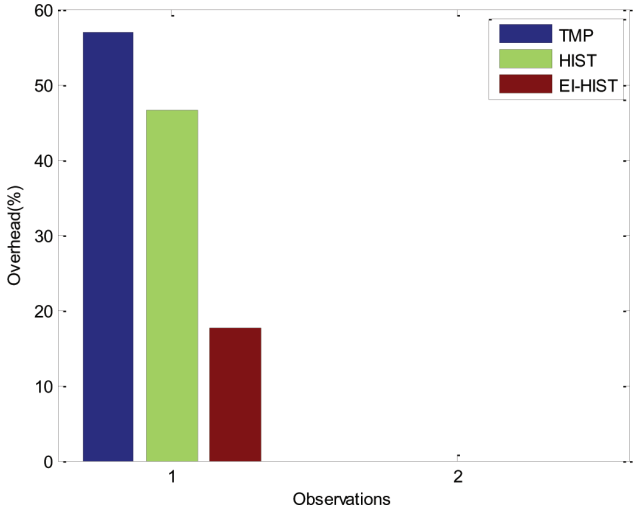

Figure 12 Overhead Observations of the developed methods.

5 Conclusion

The efficiency of template based video compression lies in the accurate definition of template region. As the wrong interpretation increases the processing overhead and delay factor, an over sampled template minimizes the accuracy

of interpolation. In this paper, an energy interpolated template coding based on inter correlative histogram is proposed. The conventional model of template match predication and histogram based coding is compared to the proposed energy interpolated histogram coding. The simulation result obtained for the developed approach tested over a traffic surveillance data illustrated a higher coding accuracy along with improvement in coding speed due to accurate template derivation.

References

[1] Wang, M., Hua, X. S., Tang, J., and Hong, R. (2009). Beyond distance measurement: constructing neighborhood similarity for video annotation. *IEEE Transactions on Multimedia,* 11(3), 465–476.

[2] Yamato, J., Ohya, J., and Ishii, K. (1992). Recognizing human action in time-sequential images using hidden markov model. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1992. *Proceedings CVPR'92*., (pp. 379–385). IEEE.

[3] Davis, J. W., and Bobick, A. F. (1997). The representation and recognition of human movement using temporal templates. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997. Proceedings.*, (pp. 928–934). IEEE.

[4] Laptev, I. (2005). On space-time interest points. *International journal of computer vision,* 64(2–3), 107–123.

[5] Dollar, P., Rabaud, V., Cottrell, G., and Belongie, S. (2005). Behavior recognition via sparse spatio-temporal features. In *2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, (pp. 65–72). IEEE.

[6] Thi, T. H., Zhang, J., Cheng, L., Wang, L., and Satoh, S. (2010). Human action recognition and localization in video using structured learning of local space-time features. In *Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, (pp. 204–211). IEEE.

[7] Ryoo, M. S., and Aggarwal, J. K. (2009). Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities. In *IEEE 12th international conference on Computer vision*, (pp. 1593–1600). IEEE.

[8] Mikolajczyk, K., and Schmid, C. (2002). An affine invariant interest point detector. In *European conference on computer vision* (pp. 128–142). Springer, Berlin,

[9] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91–110.

[10] Scovanner, P., Ali, S., and Shah, M. (2007). A 3-dimensional sift descriptor and its application to action recognition. In *Proceedings of the 15th ACM international conference on Multimedia* (pp. 357–360). ACM.

[11] Laptev, I., Marszalek, M., Schmid, C., and Rozenfeld, B. (2008). Learning realistic human actions from movies. In *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008.* (pp. 1–8). IEEE.

[12] Kläser, A., Marszalek, M., and Schmid, C. (2008). A spatio-temporal descriptor based on 3d-gradients. In *BMVC 2008-19th British Machine Vision Conference* (pp. 275–1). British Machine Vision Association.

[13] Shao, L., Jones, S., and Li, X. (2014). Efficient search and localization of human actions in video databases. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(3), 504–512.

[14] Zepeda, J., Turkan, M., and Thoreau, D. (2015). Block prediction using approximate template matching. In *23rd European Signal Processing Conference (EUSIPCO)*, (pp. 96–100). IEEE.

[15] Chen, T., Sun, X., and Wu, F. (2010). Predictive patch matching for inter-frame coding. In *Visual Communications and Image Processing*, (Vol. 7744, p. 774412). International Society for Optics and Photonics.

Biographies

Shivprasad P Patil received his B.E degree in Electronics Engineering from University of Pune, India, in 1989 and Master degree from Swami Ramanad Tirth Marathwada University, Nanded, in 2000. He is working as a Professor in the department of Information Technology in NBN Sinhgad School of Engineering, Pune. He is pursuing his PhD from Aarhus University, Denmark. He has published 06 papers in international journals and conference

proceedings in US and India. His research interests are in the areas of computer vision, multimedia data analysis and wireless multimedia communications.

Rajarshi Sanyal is a telecommunication network architect at BICS (Belgacom International Carrier Services) based in Brussels, Belgium. He is one of those leading innovation and architectural initiatives in the domain of 5G, IoT, VoLTE, to name a few. He has 20 years of experience in engineering, design, research and development in the area of wireless networks. He had obtained Ph.D. from Aalborg University, Denmark on future mobile networks. He has 5 patents and 28 research publications in the field of telecom.

Ramjee Prasad is a Professor of Future Technologies for Business Ecosystem Innovation (FT4B1) in the Department of Business Development and Technology, Aarhus University, Denmark. He is the Founder President of the CTIF Global Capsule (CGC). He is also the Founder Chairman of the Global ICT Standardisation Forum for India, established in 2009. GISFI has the purpose of increasing of the collaboration between European, Indian, Japanese, North-American and other worldwide standardization activities in the area of Information and Communication Technology (ICT) and related application areas.

He has been honored by the University of Rome "Tor Vergata", Italy as a Distinguished Professor of the Department of Clinical Sciences and Translational Medicine on March 15, 2016. He is Honorary Professor of

University of Cape Town, South Africa, and University of KwaZulu-Natal, South Africa.

He has received Ridderkorset of Dannebrogordenen (Knight of the Dannebrog) in 2010 from the Danish Queen for the internationalization of top-class telecommunication research and education.

He has received several international awards such as: IEEE Communications Society Wireless Communications Technical Committee Recognition Award in 2003 for making contribution in the field of "Personal, Wireless and Mobile Systems and Networks". Telenor's Research Award in 2005 for impressive merits both academic and organizational within the field of wireless and personal communication, 2014 IEEE AESS Outstanding Organizational Leadership Award for: "Organizational Leadership in developing and globalizing the CTIF (Center for TeleInFrastruktur) Research Network", and so on.

He has been Project Coordinator of several EC projects namely, MAGNET, MAGNET Beyond, eWALL and so on.

He has published more than 30 books, 1000 plus journal and conference publications, more than IS patents, over 100 Ph.D. Graduates and larger number of Masters (over 250). Several of his students are today worldwide telecommunication leaders themselves.