Turkish Journal of Computer and Mathematics Education *Vol.10 No.02 (2019), 946-953*
DOI: https://doi.org/10.17762/turcomat.v10i2.13574

Research Article

Designing Of Efficient Model for Facial Feature Identification and Classification Using a Sparse Fingerprint Algorithm

Sonali Gupta
Department of Computer Science & Information Technology, Graphic Era Hill University, Dehradun Uttarakhand India 248002

Abstract
By leveraging the geometrical and textural data of the retrieved local characteristics to align the probing partial face to gallery faces throughout this process, we present a novel partial face identification technique. Our fundamental assumption is that the cost function of our alignment approach ought to be reduced if the probing partial face patch and the gallery face picture are using the exact same individual. The similarity between the partial probing patch also the gallery photos across the identified facial feature points is also computed using a point-set distance metric that we provide. The usefulness of the suggested technique is demonstrated by experimental findings on four popular face datasets. This method focuses on choosing localised features from facial detection photos to distinguish between classes depending upon regression results, or partial F-test. The results demonstrate that traditional procedures are more resilient in terms of appropriate feature selection and categorization. The most noticeable characteristics were chosen by introducing a reliable method termed stepwise linear discriminant analysis, which concentrates on choosing the localised features from the face frames and classifying them relied upon regression values. A feature extraction procedure's goal is to extract the localised characteristics from faces that the earlier feature extraction methods were unable to analyse.

1. Introduction

This study examines the use of feed-forward neural networks (classifier) and Gabor filters to extract features from still images of the human face in order to recognise seven distinct facial emotions. This study provides a straightforward way for identifying facial expressions. Due to the lack of efficient expression recognition algorithms, the potential applications of facial expression recognition in various facets of daily life are still being completely realised. This research explores the use of feed-forward neural networks (classifier) and Gabor filters to extract features from still images of the human face in order to recognise seven distinct facial emotions. This study provides a straightforward way for identifying facial expressions.

So far, there have been two main methods for computer vision-based face expression detection: 2D still picture recognition and image sequence recognition. Picture sequence-based methods frequently analyse the image sequence optically before employing tools for pattern recognition to identify optical flow patterns connected to certain face expressions. This method's real-time speed and robustness are constrained since it has to acquire numerous frames of pictures in order to detect facial emotions. The challenge with this strategy is developing a feature extraction methodology that works effectively despite variations in human subjects and environmental factors. Facial expression recognition using still photos frequently uses feature-based approaches for recognition and hence has pretty rapid performance.

Frame-based approaches aim to recognise facial expressions from a single picture (Usually, the peak of the phrase). A dynamic facial event, on the other hand, changes with time from the beginning to the end. Consequently, it is easier to understand face emotions in videos. Although some frame-based algorithms capture the dynamics of facial emotions using features culled from many frames, Dynamic classification frameworks provide a more rational method of achieving this. With some noteworthy omissions, the majority of dynamic techniques to facial expression categorization are built using iterations of Dynamic Bayesian Networks, such as Hid-den Markov Models and Conditional Random Fields. For instance, categorization of emotions was achieved by contrasting the probabilities of separate HMMs trained for each category of emotions. The categorization of expressions has, however, been demonstrated to benefit more from discriminative models based on CRFs. A Hidden Conditional Random Field, which is an extension of the linear-chain CRF paradigm, is one example where the temporal dynamics of facial expressions are modelled using a further layer of (hidden) variables.

The framework was trained using picture sequences, however the categorization of the expressions was carried out by choosing by far the probable class (specifically, the emotion type) in each case. The researchers demonstrated that the approach had greater discriminatory power than the conventional linear-chain CRF when there is an additional layer of hidden variables, and (ii) modelling the temporal unfolding of the facial shapes is more crucial for facial expression exclusion than their spatial variation (considering contrasts with SVMs as a basis). Another variation of H-CRF, known as partially-observed H-CRF, was suggested in which more hidden variables are introduced to the model to encode the presence of subsets of AU pairings in each picture frame, and which are considered throughout studying, to be aware of. The regular H-CRF, which does not take into account any prior knowledge about the AU co-currencies, was outperformed by our technique. These models are in contrast to the Hidden Conditional Ordinal Random Field frameworks, which represent the ordinal correlations among the temporal phases of emotion. These models fared better than nominal H-CRF models, which do not place order constraints on their latent states. The fundamental drawback of the aforementioned designs have the ability to only have nominal or ordinal latent states, not both.

The tagging of temporal sequences presents a wide range of challenges in face and gesture research. In this article, a discriminative model for these sequence labelling tasks is introduced. Latent dynamics are split into two layers in this model, each having distinct roles. The neural network or gating layer, which is the first layer, seeks to extract non-linear correlations among the input data as well as the output labels. The second layer, the hidden-states layer, uses learning of hidden states and associated transition dynamics to simulate the temporal substructure in the sequence. For this model's training, a brand-new regularisation term is suggested that promotes hidden-state variety. We assess our model's performance on an audiovisual dataset for emotion identification also contrasting with other well-liked techniques for sequence labelling.

The analysis of facial emotions and action units is smoothly coupled and done concurrently using a unique graphical model technique, which is described in this study. Our approach is depending upon hidden conditional random fields, in which the hidden variables are related to the picture frame-by-frame action units, and also the output class label is coupled to the underlying emotion of a sequence of facial expressions. Since HCRFs are only formed via clique constraints, their labelling for hidden variables frequently lacks a logical and significant configuration. By incorporating a partially observed HCRF model, we are able to fix this issue and, in order to get over the training challenges it causes, we construct an effective plan using the Bethe energy approximation. We also suggest an online approach for real-time applications that will enable accurate incremental inference to be performed.

Increased use of hidden variables or the number of potential hidden states can both produce the desired result. In our tests, we find that applying HCRFs to image sequences with comparable appearances might result in rather diverse hidden-state configurations. While identifying the underlying emotion is our aim, finding the action units in each frame of the image proves to be essential for making the forecast.

2. Literature Survey

We utilise the sum-product network to mimic different BoW topologies. SPN is a combination of bags of words (BoWs) with an exponentially huge number of mixing components, where smaller components are reused by bigger ones. This works well for recording different activity structures as needed. Our findings show that the combined restrictions of codeword counts across the video may be modelled in a way that is expressive enough to describe the stochastic structure of somewhat complicated activities. We have dealt with the localisation and detection of actions with stochastic structure.Parsing the SPN graph, which results in the most likely explanation for the video, is what SPN inference entails. On the benchmark and or volleyball datasets, classification accuracy, localization precision, also recall of suggested strategy outperform those that are cutting-edge [1].

To quickly pinpoint the area of a video that would increase a classifier's score, the authors provide an effective method that makes use of top-down activity information. We create a 3D graph from an unique movie, where nodes represent local video subregions, and connection between nodes is based on spatial and temporal closeness. This implies that we can quickly determine the sequence's spatial and temporal regions that best match a learnt activity model. The suggested method has a number of significant characteristics. First, the detection approach is equal to a thorough sliding window search is conducted in the particular case when space-time nodes are discrete video frames. Second, demonstrate how to construct more generic versions of the network that enable "non-cubic" detection zones and possibly consent time-hopping spanning ineffective frames that may presumably deceive the classifier. Improved accuracy is the final result. The approach also supports as it is adaptable as a general tool for activity detection thanks to a wide family of characteristics and classifiers. At contrast to ST-Cube-Subvolume, which may become imprisoned in cube-shaped maxima, Outcomes shows how the location of activity changes over time and how effectively suggested space-time node structure handles this.

An innovative branch-and-cut subgraph structure for detecting activity is presented by the authors. It considerably cuts down on computation time when compared to the standard sliding window search. Because of its adaptable node structure, it allows more reliable detection of background noise. Additionally, our innovative high-level descriptor for complicated tasks shows potential. The suggested description increases accuracy for 5 of the 8 verbs. It decreases accuracy for other verbs that have a wide range of objects since the object detector sometimes fails. The spatio-temporal linkages amongst people and things in the video can be preserved even though taking use of the quick subgraph search thanks to our innovative high-level descriptor, which also offers potential for challenging tasks [2].

In this study, the author aims to understand how important they are for supporting efficient inference in realistic films, as well as what activity components and their spatiotemporal relationships could perhaps be depicted to simulate complicated human activities. This builds on earlier research that often doesn't consider the "what" question. Our learning has two objectives. We get knowledge of the activity model's structure as well as the PDFs connected to its nodes and edges. Then, using this model, fresh films are parsed in order to localise activity portions that are pertinent and present at various sizes. Our inference is made arbitrary permutations of nodes in spatial temporal graphs invariant, accounting for any mistakes in extracting video tubes. The same goal used for video processing is used to train our activity model. We learn the archetypal graph and pdfs derived from a collection of training spatiotemporal graphs and connected to the design nodes and edges. Both inference and learning are expressed in terms of a resilient, least-squares optimization that is unaffected by random permutations of the nodes in spatiotemporal networks. In order to find and localise pertinent activity portions in fresh movies, the model is employed for parsing. On standard datasets from the Olympic and UT human interactions, suggested strategy exceeds the latest technology. Within the same framework—that of weighted least-squares—the author has developed inference and learning of a structural activity model. In this innovative volumetric-based method to activity identification and video parsing, we present: [3].

The use of CNNs for human action recognition in videos is discussed by the author in this work. And demonstrate how various characteristics may be recovered from the input by doing several different convolutional processes at the same time. A further benefit of CNN-based models is their feed-forward nature, which makes the recognition phase extremely effective. The majority of previous work depends on subject expertise to create intricate handmade features from inputs. Deep models called Convolutional neural networks automate the development of features by acting directly on the raw inputs. In this research, the authors create a brand-new 3D CNN action recognition model. This model uses 3D convolutions to extract characteristics from the spatial and temporal dimensions [4].

It is not necessary to explicitly segment the whole video clip in order to perform action recognition as temporal event detection. With the help of a multi-class SVM that optimises the gap between classes, we train our discriminative recognition model using labelled data. Once every action's model has been learnt, dynamic programming is an effective method for doing segmentation and recognition simultaneously. For the purpose of simultaneously segmenting time and identifying actions in video, we suggested a unique method. Dynamic programming was used to effectively perform segmentation inference while multi-class SVM was used to train the recognition model discriminatorily. The advantages of our strategy over cutting-edge techniques are demonstrated by experimental findings on the honeybee, Weizmann, and Hollywood datasets [5].

Data parallelism is supported via the use of asynchronous SGD. On a cluster of 1,000 computers (16,000 cores), the model was trained in a distributed manner over the course of three days. It is feasible to create high-level features from unlabeled data, according to experimental findings employing classification and visualisation techniques. Our network was trained to recognise 22,000 different object categories from ImageNet with 15.8% accuracy, a 70% relative increase over the prior state-of-the-art. By employing many clones of the main model and an asynchronous SGD, we increased the training scale even more. Through a collection of central "parameter servers," the models exchange updated information. Compared to traditional (synchronous) SGD, asynchronous SGD is more resilient to errors and slowness. The transmission of gradients and parameters between the parameter servers and the model partitions is automatically handled by our DistBelief architecture. Every SGD step in our training involves computing the gradient on a minibatch of 100 instances. For further information on how we applied the optimization, see Appendices B, C, and D. In order to learn invariances from unlabeled data, we combined concepts from newly created algorithms [6].

3. Proposed System

A novel partial face recognition method, known as the Sparse Fingerprint Classification Algorithm, is suggested in order to overcome this partial face identification issue (SFCA). Multi-Directional Multi-Level Dual Cross Patterns is a novel method that was suggested in order to prevent the quality of the facial picture from degrading and significant fluctuations caused by lighting, position, occlusion, and expression. The histogram-based face description approaches that break the face into tiny blocks, furthermore uniformly apply sample codes are extracted in this suggested method, making it superior to other current methods. The grid is then categorised likewise sampling expression-related data at various scales to create the face descriptor.

The concept behind dimension reduction by identifying discriminating characteristics is to maximise overall data dispersion while limiting variation within classes. It is clear that there is a significant degree of feature value fusion across the six classes, which may lead to a high percentage of misclassification. It should be noted that there may really be more elements than just the first three; nonetheless, these three were chosen to produce the best visual effect. The work uses a strong characteristic as a result. This is simpler to understand than other approaches now in use, has strong predictability, and requires less computing power. We deliberately use full covariance Gaussian distributions in the feature functions at the observation level in order to overcome this restriction.

Numerous methods have been employed for precise expression categorization in the classification module. The recognition rate for the numerous facial expressions that the authors were able to identify using artificial neural networks (ANNs) was 73%. The explicit categorization of potential basic linkages by ANN, however, is a black box with limited capabilities. In addition, ANNs may require a lot of training time and may become trapped in a poor local minima. Additionally, the authors' FER system made use of support vector machines (SVMs). SVMs, however, do not use direct estimate of the probability; instead, the observation probability is derived via indirect methods. Furthermore, SVMs only ignore the temporal correlations between video frames, hence it is assumed that every frame will be statistically independent of the others.

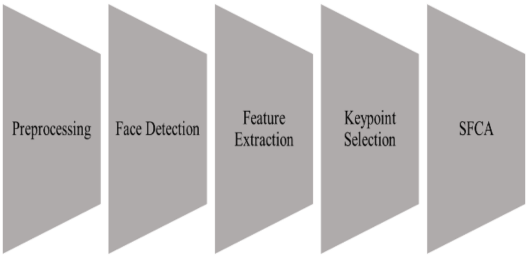

Fig 1: Block Diagram

Three separate tests were conducted to demonstrate the efficacy of SH-component FER's parts. According to the recognition rate, the best scenario from the first experiment was chosen for this purpose. Thereafter, three related experiments utilising the 10-fold validation criteria were carried out. In the first instance, ICA rather than SWLDA was used with HCRF. In the second instance, ICA and LDA were combined before the features were sent to HCRF. The following is a list of the proposed system's several benefits:

- The SIFT extraction increases the feature's stability.
- The use of an efficient classification approach increases the recognition rate.

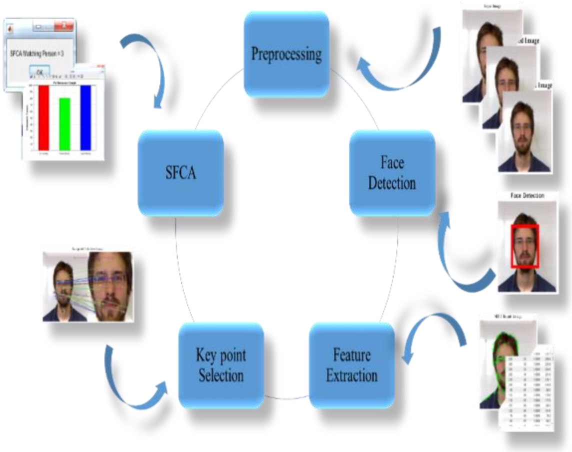

Fig 2: System Architecture

4. Results

By leveraging the geometrical and textural data of the retrieved local characteristics, we propose a new partial face identification method in this process, matching the probing partial face to gallery faces. According to our fundamental logic, our alignment method's cost function has to be scaled back if the probing partial face patch as well as the gallery face picture are from the same individual. We also provide a point-set distance metric to calculate the similarity between the partial probing patch and the gallery photos across the identified facial feature points. Practical results on four well-known face datasets show the effectiveness of the recommended approach. This method focuses on choosing localised features from facial detection photos to distinguish between classes relied upon regression results, or partial F-test. The results demonstrate that traditional procedures are more resilient in terms of appropriate feature selection and categorization. Sequential linear discriminant analysis, which emphasises selecting localised traits, from the face frames and classifying them depending upon regression values, was proposed as a reliable method for choosing the most noticeable features. A feature extraction methodology's goal is to extract the localised characteristics from faces that the earlier feature extraction methods were unable to analyse.

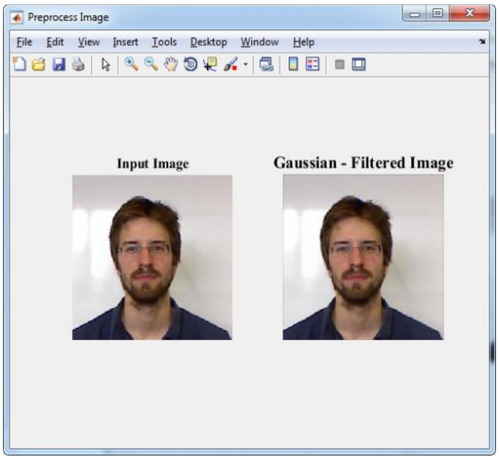
Preprocess Image
File Edit View Insert Tools Desktop Window Help
Input Image
Gaussian - Filtered Image

Fig 3: Guassian Filtered Image

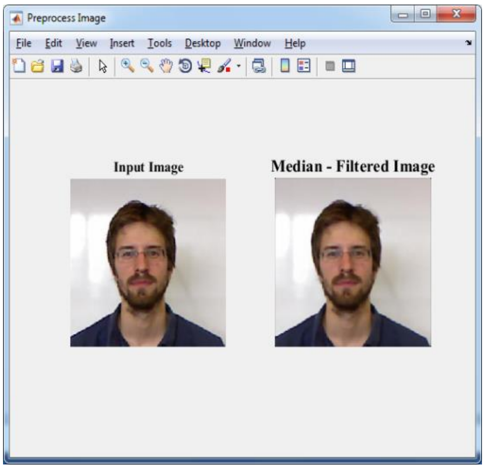

Fig 4: Median Filtered Image

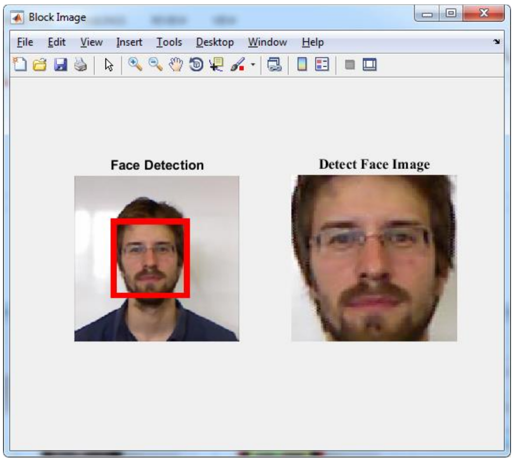

Fig 5: Face Detection

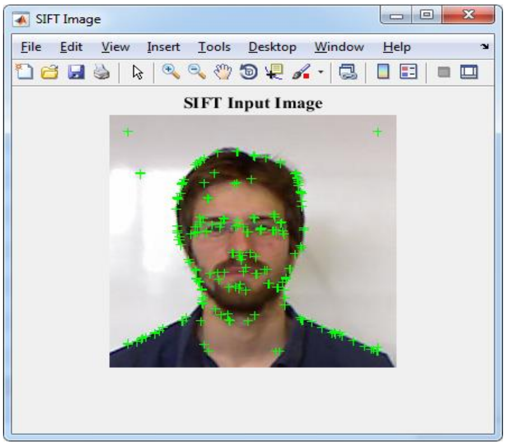

Fig 6: SIFT Input Image

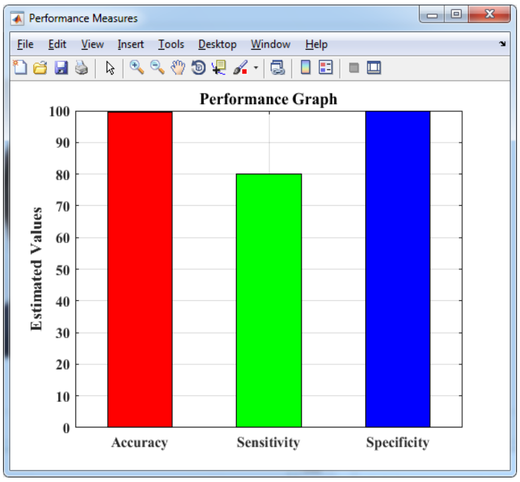

Fig 7: Performance Analysis

5. Conclusion

In this research, we've suggested a comprehensive feature set matching approach for partial face recognition. The suggested RPSM approach can match the probing partial face to gallery facial photos reliably regardless of whether an obstruction is present, random partial crop, and extreme facial emotions. Following face alignment, partial face recognition is accomplished through calculating the degree of similarity between faces based on the suggested point set distance, which is easily obtained with the face alignment result. The robust matching algorithm used by the SFCA, which takes into account both textural and geometrical similarities, is its distinguishing feature.

6. Future Enhancement

In this part, we show the suggested method for classifying textures. The RLBP operator, which was created by rotating the weights of the LBP operator, is introduced first. Additionally, the inherent structure of the patterns is used by applying the uniform pattern principle to produce uniform RLBP (uRLBP). The direction with the largest difference in the circular neighbourhood is referred to as the dominant direction. When an image rotates, the predominant direction in a neighbourhood likewise rotates by the same angle.

Reference

[1] M. H. Siddiqi, F. Farooq, and S. Lee, “A robust feature extraction method for human facial expressions recognition systems,” in Proc. 27th Conf. Image Vis. Comput. New Zealand , 2012, pp. 464–468.

[2] M. H. Siddiqi, A. M. Khan, T. C. Chung, and S. Lee, “A precise recognition model for human facial expression recognition system,” in Proc. 26th IEEE Can. Conf. Elect. Comput. Eng., 2013.

[3] Li, W.J.; Wang, J.; Huang, Z.H.; Zhang, T.; Du, D.K. LBP-like feature based on Gabor wavelets for face recognition. Int. J. Wavelets Multiresolution Inf. Process. 2017, 15, 1750049.

[4] Masi, I.; Wu, Y.; Hassner, T.; Natarajan, P. Deep Face Recognition: A Survey. In Proceedings of the 31st SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI 2018, Parana, Brazil, 29 Octorber–1 November 2018; pp. 471–478.

[5] V. Bettadapura. (2012). “Face expression recognition and analysis: The state of the art.” [Online]. Available: http://arxiv.org/abs/1203.6722

[6] C. Lisetti and C. LeRouge, “Affective computing in tele-home health: Design science possibilities in recognition of adoption and diffusion issues,” in Proc. 37th IEEE Hawaii Int. Conf. Syst. Sci., Hawaii, USA, Jan. 2004, pp. 348–363.

[7] X. Wu and J. Zhao, “Curvelet feature extraction for face recognition and facial expression recognition,” in Proc. 6th Int. Conf. Natural Comput. (ICNC) , vol. 3. Aug. 2010, pp. 1212–1216.

[8] S. Moore and R. Bowden, “The effects of pose on facial expression recognition,” in Proc. Brit. Mach. Vis. Conf., 2009, pp. 1–11.

[9] Z. Zhu and Q. Ji, "Robust real-time face pose and facial expression recovery," in Proc. Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 1. Jun. 2006, pp. 681–688.

[10] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: Machine learning and application to spontaneous behavior," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 2. Jun. 2005, pp. 568–573.

[11] Y.-L. Tian, T. Kanade, and J. F. Cohn, "Facial expression analysis," in Handbook of Face Recognition. New York, NY, USA: Springer-Verlag, 2005, pp. 247–275.

[12] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," Image Vis. Comput., vol. 27, no. 6, pp. 803–816, 2009.

[13] C. Shan and R. Braspenning, "Recognizing facial expressions auto-matically from video," in Handbook of Ambient Intelligence and Smart Environments. New York, NY, USA: Springer-Verlag, 2010, pp. 479–509.

[14] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Toward enhanced P300 speller performance," J. Neurosci. Methods, vol. 167, no. 1, pp. 15–21, 2008.

[15] M. Ager, Z. Cvetkovic, and P. Sollich. (2013). "Phoneme classifica-tion in high-dimensional linear feat ure domains." [Online]. Available: http://http://arxiv.org/abs/1312.6849

[16] M. Nusseck, D. W. Cunningham, C. Wallraven, and H. H. Bülthoff, "The contribution of different facial regions to the recognition of conversational expressions," J. Vis., vol. 8, no. 8, 2008, Art. ID 1.

[17] Sayed, A., Sardeshmukh, M., & Limkar, S. (2014). Improved Iris Recognition Using Eigen Values for Feature Extraction for Off Gaze Images. In ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India-Vol II: Hosted by CSI Vishakapatnam Chapter (pp. 181-189). Springer International Publishing.

[18] R. K. Jha, S. V. Limkar, and U. Dalal. 2011. A performance comparison for QoS with different WiMAX environment for video application. In Proceedings of the International Conference &amp; Workshop on Emerging Trends in Technology (ICWET '11). Association for Computing Machinery, New York, NY, USA, 785–790. https://doi.org/10.1145/1980022.1980194