Spatial Frequency based Video Stream Analysis for Object Classification and Recognition in Clouds

Muhammad Usman Yaseen
College of Engineering and Technology
University of Derby
Derby, Uk
M.Yaseen@derby.ac.uk

Ashiq Anjum
College of Engineering and Technology
University of Derby
Derby, Uk
A.Anjum@derby.ac.uk

Nick Antonopoulos
College of Engineering and Technology
University of Derby
Derby, Uk
N.Antonopoulos@derby.ac.uk

ABSTRACT

The recent rise in multimedia technology has made it easier to perform a number of tasks. One of these tasks is monitoring where cheap cameras are producing large amount of video data. This video data is then processed for object classification to extract useful information. However, the video data obtained by these cheap cameras is often of low quality and results in blur video content. Moreover, various illumination effects caused by lightning conditions also degrade the video quality. These effects present severe challenges for object classification. We present a cloud-based blur and illumination invariant approach for object classification from images and video data. The bi-dimensional empirical mode decomposition (BEMD) has been adopted to decompose a video frame into intrinsic mode functions (IMFs). These IMFs further undergo to first order Reisz transform to generate monogenic video frames. The analysis of each IMF has been carried out by observing its local properties (amplitude, phase and orientation) generated from each monogenic video frame. We propose a stack based hierarchy of local pattern features generated from the amplitudes of each IMF which results in blur and illumination invariant object classification. The extensive experimentation on video streams as well as publically available image datasets reveals that our system achieves high accuracy from 0.97 to 0.91 for increasing Gaussian blur ranging from 0.5 to 5 and outperforms state of the art techniques under uncontrolled conditions. The system also proved to be scalable with high throughput when tested on a number of video streams using cloud infrastructure.

Keywords

Empirical Mode Decomposition; Local Ternary Patterns; Riesz Transform; Amplitude Spectrum, Cloud Computing, Big Data Analytics, Object Classification

BDCAT'16, December 06-09, 2016, Shanghai, China

DOI: http://dx.doi.org/10.1145/3006299.3006322

1. INTRODUCTION

With the advancements in multimedia technology, it is now becoming easier to offer large scale monitoring of critical events. The monitoring cameras are capable of recording and storing activities within these events. The video data generated by these monitoring cameras is processed manually and automatically to extract useful information such as person identification and tracking. However, most of the cameras used for monitoring are cheap off the shelf cameras because of budget constraints. The video streams generated by these cameras often contain blur frames due to motion, lack of focus, or atmospheric turbulence. Since these cameras work under uncontrolled lighting conditions they are also prone to illumination effects. Rotation angle of the objects being monitored is also another challenge posed by these cameras.

The accuracy of any object classification system is highly dependent on how these challenges are tackled. A good countermeasure to these challenges can lead to highly accurate results. However, video de-blurring and de-illumination are resource and time consuming tasks and often bring in new artifacts [1]. It is therefore desirable to perform classification with a process that is invariant to blur and illumination. Various approaches have been proposed in the past to tackle these problems. The most prominent among these approaches are built on top of insensitive moments [2], color constancy [3] and Fourier phase. However, these approaches are designed to perform classification globally and do not take into account local properties of objects. Various solutions based on the magnitude, phase and spectral information [4] have also been designed, however, these methods focused on texture analysis whereas blur and illumination invariance was not considered as a design criterion.

One of the successful approaches to address these challenges is to perform analysis of image or video frames by shifting them from spatial domain to spatial-frequency domain. In spatial domain, processing of video frames is performed by directly using the gray values of pixels. Spatial frequency domain allows the processing of video frames by projecting them on a set of basis functions which are defined by the method itself. This phenomenon expands the video frame into frequency components with both high and low magnitudes. Variety of methods exist for the conversion of spatial domain signal to spatial frequency domain. These methods include Fourier Transform [5], Wavelet Transform [6], and Wigner distribution [7]. However, these methods are not

adaptive.
Huang et al. [8] proposed a method known as Empirical Mode Decomposition (EMD) for image analysis. EMD expands a signal into its frequency components adaptively. These frequency components are termed as Intrinsic Mode Functions (IMFs). EMD tries to extract highest frequency components from the original input signal in each mode. It separates the locally highest frequencies and stores them into an IMF. The rest of the IMFs contain the remaining frequencies in the lowest order which ends up in a residual part. In order to apply empirical mode decomposition on the images, two dimensional empirical mode decomposition (2DEMD) or Bi-dimensional empirical mode decomposition (BEMD) was introduced [9].
We have used Bi-dimensional empirical mode decomposition to decompose a video frame into Intrinsic Mode Functions in this paper. A video frame is first extracted from video stream and the BEMD is applied to decompose it into IMFs. BEMD provides several advantages over spatial domain analysis. Features can be extracted easily according to the distribution of local phase or energy. Each IMF is analyzed independently and in parallel using Reisz transform to extract local properties (amplitude, orientation, and phase) of a video frame. These local properties are further examined to perform classification tasks using local pattern features. As the process is data intensive, a cloud infrastructure has been utilized to meet the challenges of scalability and performance and to process the data quickly and efficiently.
The main contributions of this paper are: Firstly, we pioneer the use of EMD on video streams in a cloud based distributed environment. We have shown that only first three IMFs are sufficient to perform classification under challenging conditions with a high accuracy rate. This is advantageous in two ways: I) Reduced feature extraction time as compared to other methods. II) Illumination and blur invariance, since only lower IMFs are sensitive to variation effects. Secondly, we utilized the amplitude property derived from each IMF using first order Reisz transform and showed that it provides high accuracy than the phase or orientation properties. Thirdly, we propose a stack based hierarchy of local pattern features generated from the amplitude of each IMF for highly accurate classification.
The organization of rest of the paper is as follows: Section II provides the review of state of the art techniques that have been used recently for object classification. Section III explains the approach and implementation of our object classification system. The experimental setup is described in Section IV. The experimental results and their analysis are detailed in Section V. Section VI concludes the paper with a glimpse of the future work.

2. RELATED WORK

Object classification has been the focus of many studies for the last several decades. However, classifying objects from video streams under uncontrolled conditions poses many challenges and is now acquiring much attention from the research community. We provide a brief review of the recent approaches proposed in this domain and identify the research gaps.
A number of authors have used color information to develop blur and illumination robust descriptors. A robust descriptor has been proposed by Joost et al. [3]. They used ratios of image derivatives to develop a descriptor for color constant ratios that is invariant to blur and illuminant color. A similar kind of approach based on color information has been proposed by Ballard et al. [10]. They made use of color histograms to perform recognition of objects. Another approach based on color histograms is proposed by Funt et al. [11]. They utilized color constant derivatives to represent an object for recognition. However, these approaches could not perform well against variations in illuminant color especially with a change in the camera viewpoint or object orientation and are also dependent on lightening geometry.
The moment invariants have also been used in the past as blur invariant features. Flusser et al. [12] pioneered the use of moment invariants developed on top of geometric moments. They also utilized central moments to provide invariance to translation. Moment invariants have also been the part of various applications such as template matching [13], recognition of defocused objects and X-ray imaging. Complex moments were also proposed [14] for blur, rotation and scale invariance. Despite their wide usage in various applications, moment invariants remained sensitive to noise and background clutter.
Invariants based on the phase of frequency spectrum obtained by Fourier transform are investigated by [15]. These invariants are also insensitive to the shift of the image. Ville et al. [16] proposed a centrally symmetric blur invariant descriptor based on phase-only spectrum of an image. The phase-only spectrum was normalized so it became insensitive to linear brightness changes as well. A similar kind of approach is adopted by Ville et al. [17] in which the phase information was calculated within a local window for every image position. The quantization of the phase of discrete Fourier transform and de-correlation of low frequency components was performed in an eight dimensional subspace. A histogram of the resulting features was used for classification of blurred texture images. However, these invariants are limited to image shifts. Also, invariance to translation has not been considered in phase based frequency spectrum invariants. We propose a blur and illumination invariant feature descriptor which provides invariance to higher blur radius and high PSNR values. Interestingly, our feature descriptor also provides good results for sharp video frames that are not blurred.

3. VIDEO ANALYSIS APPROACH

We present here the approach behind our object classification system. The video streams are first acquired by the video capturing cameras and are then stored in cloud storage. The cloud manager fetches these video streams from the cloud storage and distributes them among various cloud nodes. The cloud manager is solely in-charge of the allocation of video streams to each cloud node and manages workload distribution among cloud nodes. The video streams are decoded to extract individual video frames. These video frames are artificially blurred with varying radius. Noise has also been added to the objects with different PSNR values. These objects are then classified by blur and illumination invariant feature descriptor. Figure 1 shows the approach of our blur and illumination invariant object classification system.

3.1 Decomposition of Video Frame

Each decoded video frame is decomposed into its frequency

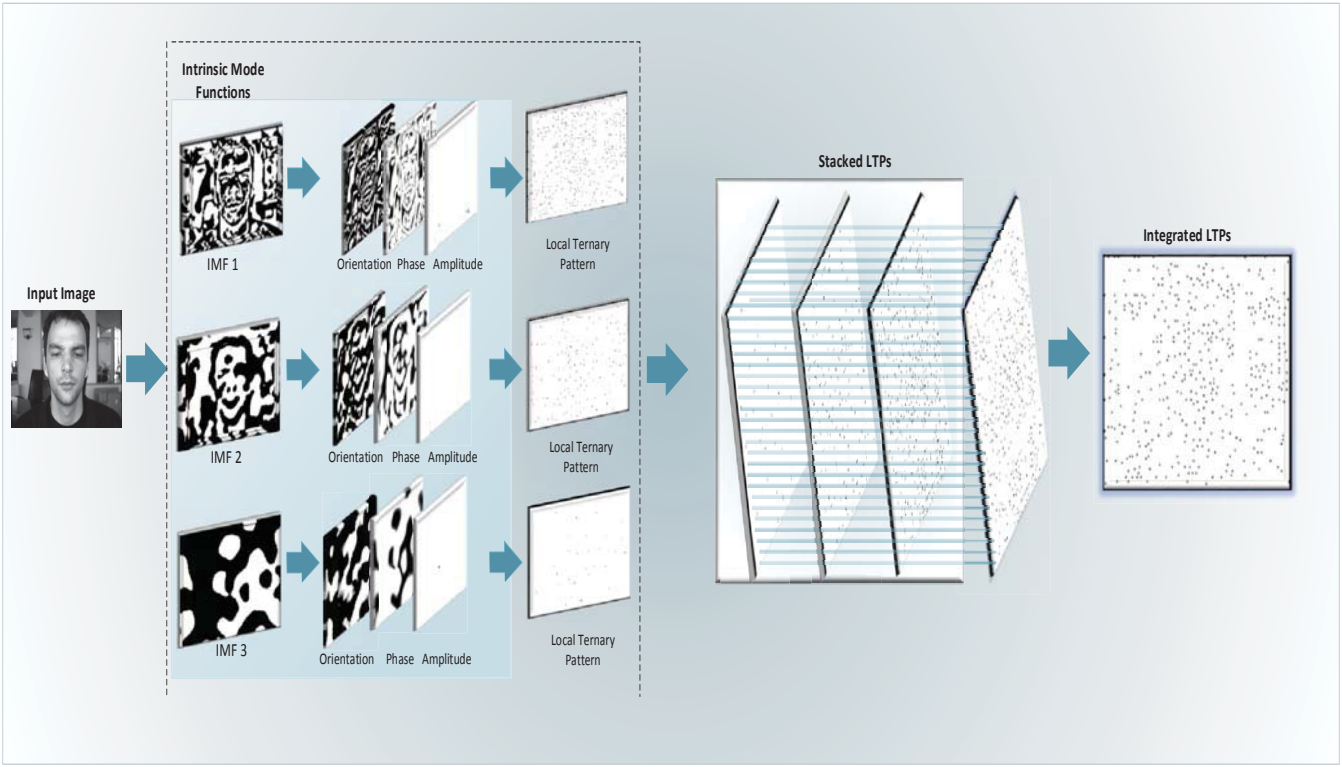

Figure 1: Video Analysis Approach

components through Two Dimensional Empirical Mode Decomposition (2DEMD) [9] which is a fully unsupervised approach. It defines its basis functions directly from the data and is not dependent on other methods. 2DEMD expands a video frame into its frequency components adaptively. These frequency components are termed as Intrinsic Mode Functions (IMFs) and are defined by the video frame itself. Sifting process [18] is used to extract these IMFs from the video frame. This algorithm tries to extract highest frequency components from the original input video frame in each mode. It separates the local high frequencies and stores them in to an IMF.
The rest of the IMFs contain the remaining frequencies in the lowest order. This ends up into the residue which contains remaining lowest frequencies. By combining all the IMFs and the residue, original signal can be obtained. EMD allows visualizing spatial-frequency characteristics signal by expanding the parent signal into IMFs as shown in figure 2. The sifting algorithm is defined as follows:

- *Extrema Identification*
  Determine the extrema points (maxima points and minima points) of the input video frame I(x, y), where I is the 2D video frame with x = 1 to M and y = 1 to N.
- *Envelop Calculation*
  Connect the extrema points (all the maxima points and the minima points respectively) using radial basis function to form lower and upper 2D envelops denoted by emax(x, y) and emin(x, y).
- *Mean Envelop Calculation*
  Average the two envelops i.e. maxima envelop and minima envelop to generate the local mean envelop m1.

$$m1(x,y) = (e_{max}(x,y) + e_{min}(x,y))/2 \tag{1}$$

- *ProtoIMF Generation*
  Subtract out the local mean from the image to generate hl.

$$hlk = I(x,y) - m1(x,y) \tag{2}$$

  Repeat the entire process until h1 is a 2IMF. The process terminates when the mean envelop is very close to zero.

Depending upon the stopping criterion, it is arbitrated that whether the mean envelop is close to zero or not. The whole procedure is reiterated if the mean envelop is not close enough to zero. The reiteration is performed a number of times until the stopping criterion is satisfied. The fulfilment of stopping criterion results in the final IMF.

$$C1(x,y) = hlk(x,y) \tag{3}$$

The residual is defined by subtracting the original input video frame from the Cl(x,y)

$$R1(x,y) = I(x,y) - C1(x,y) \tag{4}$$

The next IMF is obtained by repeating the entire procedure on the residual by considering it as an input image.

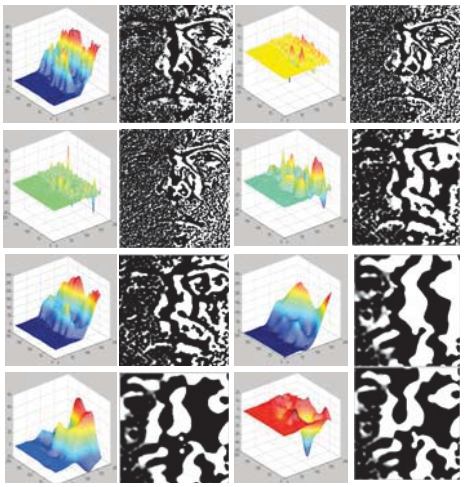

Figure 2: Averaged extrema surfaces along with their visual representation

Figure 3: Orientation, Phase and Amplitude of first three IMFs

$$I(x,y) = R1(x,y) \tag{5}$$

For all the subsequent residuals, the process is repeated to obtain various IMFs in the descending order of their frequencies. The procedure is normally stopped when there are no more extrema points in the residual frame. A video frame can be expressed as the sum of all IMFs with a residual given below:

$$I(x,y) = R1(x,y) + \sum_{i=1}^{L} C_i(x,y) \tag{6}$$

3.2 Amplitude, Phase and Orientation Spectrum

Riesz Transform [19] is applied on the IMFs afterwards to obtain monogenic data which helps to study the local properties of video frames. Monogenic data is a local quantitative and qualitative measure of the video frame. Local amplitude, phase and direction can be calculated from the monogenic signal of each IMF as shown in figure 3.

The Riesz transformed signal in the frequency domain can be expressed as;

$$F_R(v) = i\frac{v}{v} \times F(v) = h_2(v) \times F(v) \tag{7}$$

Where h_2 is the transfer function and the generalization of Hilbert transform.

The corresponding spatial domain representation can be given as;

$$F_R(x) = i\frac{x}{2\pi}x^3 \times F(x) = h_2(x) \times F(x) \tag{8}$$

The 2D analytical or monogenic signal which is constituted by the original signal and its Riesz transform is given as;

$$F_M(x) = F(x) - (i,j) \times F_R(x) \tag{9}$$

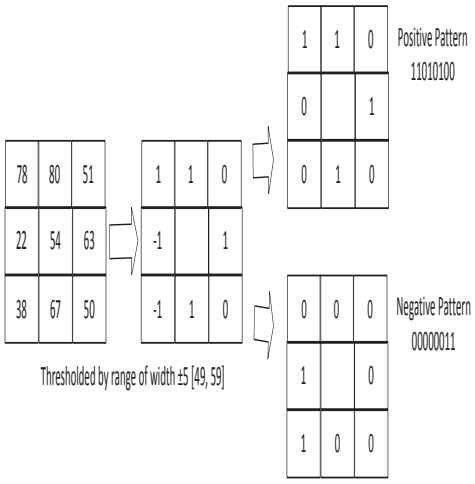

Figure 4: Generation of LTP and its split-up into positive and negative LBP codes

3.3 Local Pattern Features

We propose the use of local ternary patterns to analyze and classify objects from the video streams. Local Ternary Pattern is used as a descriptor to extract local features from intrinsic mode functions. It is an extension of local binary pattern histogram, however, it is more robust to noise as it codes the input video frames into ternary patterns instead of binary patterns like in LBP. Local ternary patterns use a threshold value to threshold the pixels into a ternary code. The pixels in the range of 't' are set to zero, the pixels above this range are set to positive 1 (+1) and the values below the range are set to negative 1 (-1) as shown in figure 4.

If 'k' is considered as threshold constant, c is taken as the value of center pixel and 'p' be the neighboring pixel then the local ternary pattern can be given as;

$$\begin{cases} 1, & \text{if } p > (c+k) \\ 0, & \text{if } p > (c-k) \ \ and \ \ p < (c+k) \\ -1, & \text{if } p < (c-k) \end{cases}$$

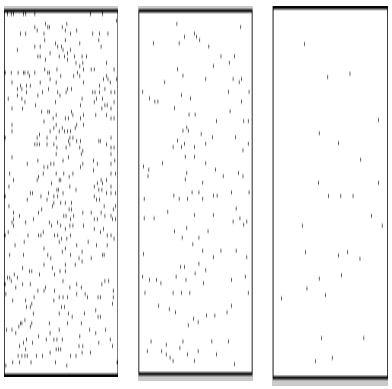

Figure 5: Local Ternary Patterns of the Amplitudes of first three IMFs

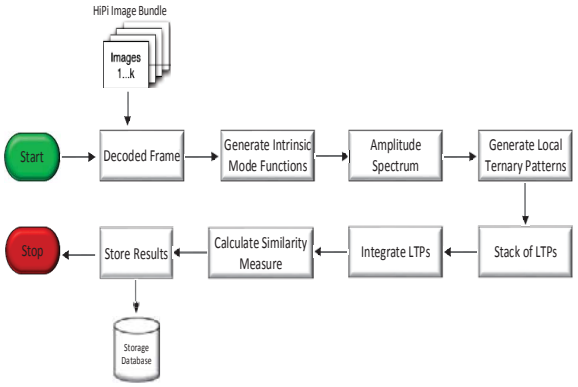

Figure 6: Workflow of the proposed system

Each threshold pixel will therefore be comprised of one of the three values. The neighboring pixels after the threshold are grouped into a three value pattern called ternary pattern. The histogram is then computed from these ternary patterns. Since the histogram results in a large range, the ternary pattern is split into two binary patterns. The resultant histogram which represents the feature descriptor is the concatenated histograms of the two binary patterns and hence the double of size of LBP.
We have calculated the local ternary patterns of the amplitudes of first three IMFs. These patterns are then arranged in a hierarchical fashion to generate feature vector. A visual representation of these LTPs is shown in figure 5.

3.4 Stack based Hierarchy of Features

We propose a stack based hierarchical approach of the local ternary patterns to represent the input video frames as a feature descriptor. The local ternary patterns generated from the amplitude spectrum of each IMF are stacked together in a hierarchical fashion such that the pixels of each IMF are adjacent to each other. All the adjacent pixels are then summed together to form an integrated local ternary pattern of the whole video frame. This represents a feature vector of the whole video frame.
LTP histogram generation is performed for all the video frames and the image which is to be matched for similarity. Matching is performed by comparing the LTP histogram of the marked object frame with all the frames of a video stream. The histogram intersection is used as a distance measure to calculate the similarity between two frames. After the face of a person is authenticated correctly, the matching score associated to it is stored in a database as depicted in Figure 6.

4. EXPERIMENTAL SETUP

This section details the experimental setup and the parameters used to evaluate the proposed system. The reported results mainly focus on the accuracy with blurred content, accuracy under various illumination conditions, and the scalability i) video stream decoding time, ii) video data transfer time to the cloud, iii) video data analysis time on the cloud nodes.
The proposed system is evaluated on an OpenStack based cloud resource. The cloud consists of six server machine. Each machine has 12 cores with two 6-core Intel Xeon processors running at 2.4 GHz with 32GB RAM and 2 Terabyte storage capacity. Each cloud instance has 72 processing cores with 192GB of RAM and 12TB of storage capacity. The experimental results reported in this paper are obtained from a 4 node cloud deployment. Each cloud node has 4 VCPUs, 8 GB RAM and data storage of 100 GB per node. Each vCPU is running at 2.4GHz.
In order to evaluate the proposed system in the cloud, Hadoop MapReduce framework [21] inspired by [26][27][28] is utilized as shown in figure 7. The empirical mode decomposition is implemented in JAVA. OpenCV [22] with JNI wrappers for the native C++ library is used as image/video processing library. Hadoop comes with Yarn which is responsible for resource management and job scheduling. The Hadoop MapReduce framework has a NameNode responsible for load balancing among the nodes. The Data/Compute Nodes are used for storing and processing the data.
The decoded frames are first bundled using the HIPI Image Bundle [23] and are then processed by map-reduce jobs. The map tasks perform the EMD and object classification while the reduce tasks collect the results. The map-reduce jobs use the configured chunk size of 128MB to define the number of input splits from the HIPI ImageBundle (HIB). Figure 7 depicts the process of video streams analysis on the cloud.

The video dataset on which empirical mode decomposition is applied is self-generated at University of Derby consisting videos of different subjects. Each video stream in the video dataset has time duration of around 120 seconds. The video streams are H.264 encoded with a resolution of 704x528 and a frame rate of 25fps. This makes a total of 3000 video frames in each video stream. The data rate and bit rate of each video stream are 421 kbps and 461 kbps.
The BioID[24] and Yale[25] Face Database are used to measure the efficiency of proposed approach. During the recording of BioID Face Database, special importance has been given to real world situations. Therefore, this database contains a diversity of face sizes, background conditions and illumination effects. The database comprises of 1521 gray level images captured at resolution of 384x286 pixels. For testing purposes, the images are artificially blurred by using a Gaussian blur mask with various sigma values (0, 0.25 . . .). Figure 8 shows some example images from BioID Face Database.

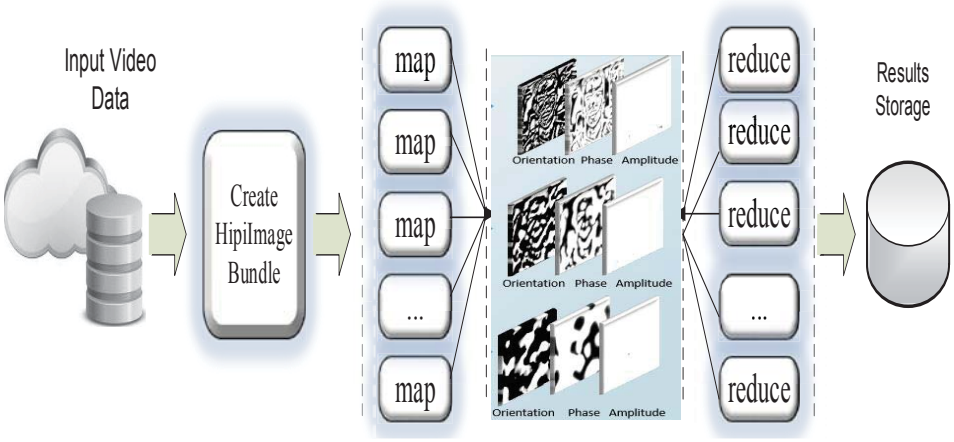

Figure 7: Video Stream Analysis on the Cloud

Figure 8: Example images from BioID Face Database with increasing artificial blur

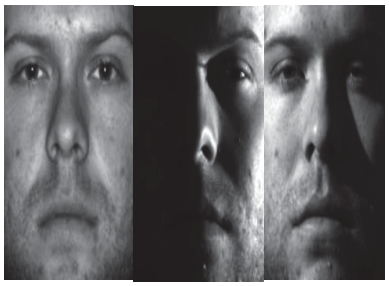

Figure 9: Example images from Yale Face Database with various illumination conditions

The Yale Face Database contains images from various subjects with different poses and illumination conditions. All the images are manually aligned and cropped having a resolution of 168x192 pixels. Every subject demonstrates variations in illumination conditions (left-right, center-right, right-right) and facial expressions (normal, sad, happy, sleepy). Figure 9 shows some example images from the Yale Face Database.

5. RESULTS AND DISCUSSION

The results obtained from the configurations described in experimental setup section are described in this section. The main focus of these results is to evaluate the system for classification accuracy with blurred content and with various varying illumination conditions. The execution of the system on the cloud evaluates the scalability and robustness of the proposed system by analyzing various components of the system such as image bundle creation time, video data transfer time to the cloud, and video data analysis time on the cloud nodes.

5.1 Accuracy with blurred content

The performance of the proposed system was evaluated with the artificially caused blurred images in the first experiment. The widely used BioID Face Database was used for this purpose. The BioID Face Database contains images which have variation in pose and expressions. These images were further artificially blurred by performing a convolution with the Gaussian blur mask. The values of the mask ranges from 0 to 5 with zero being no blur and 5 as the maximum blur. It was therefore possible to observe the joint effect of pose, expressions and blur. The mean recognition rates of the proposed system and the widely used state of the art techniques including LBP, LTP and LPQ are plotted in figure 10. It can be observed that the proposed system performs better than the existing techniques even with the minimum blur to maximum blur. The existing approaches tolerate slight blur but as the value of sigma increases, the accuracy rate falls down rapidly. On the other hand, the proposed system handles increasing blur expressively well. This is because of the fact, that blur effect mostly resides in the low frequency band. Since we are dealing with only highest three IMFs, the remaining low frequency components and the residue are automatically discarded. This helps to achieve high accuracy even with high sigma value.

5.2 Accuracy under various illumination conditions

The performance of the proposed system was evaluated with varying illumination conditions in the second experiment. The widely used Yale Face Database was used for this purpose. The Yale Face Database contains images which have variation in pose, expressions and illumination conditions. It contains images with lightning effects from dif-

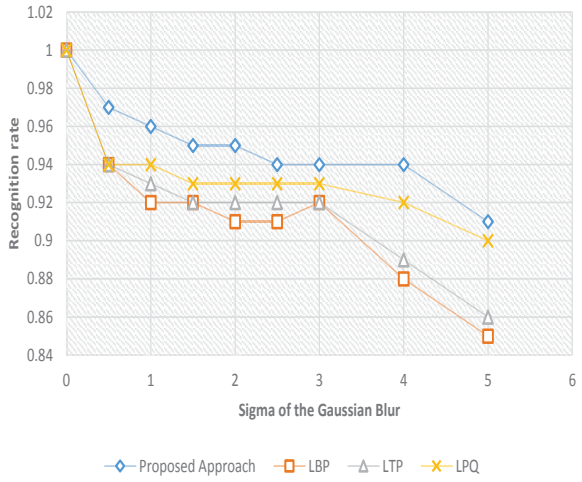

Figure 10: Mean recognition rates for increasing Gaussian blur

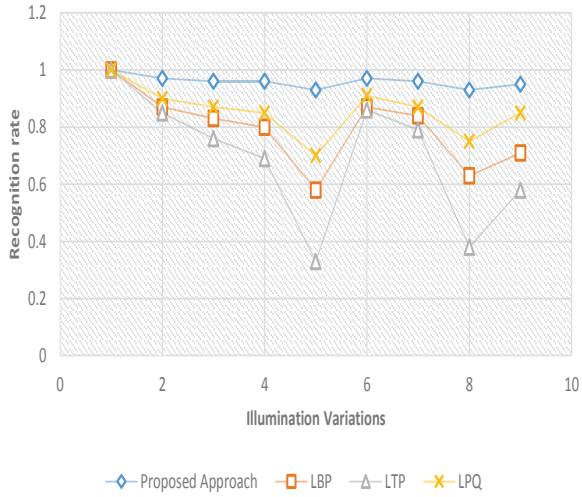

Figure 11: Mean recognition rates for various illumination conditions

ferent angles such as left-right, center-right and right-right. In-addition, the facial expressions vary with normal, happy, sad and sleepy. The mean recognition rates of the proposed system and the existing techniques are plotted in figure 11. It can be seen that the proposed system tackles various illumination conditions better than the existing techniques and maintains a smooth accuracy plot.

5.3 Video data analysis on the cloud nodes

In order to process the video frames in parallel, we have utilized the Hadoop MapReduce framework. The input data which is to be processed is first transferred to Hadoop file storage (HDFS). The results generated by the framework are stored in the database. The cloud nodes are responsible to execute analysis tasks . The map task in our system is responsible to execute empirical mode decomposition and the stack based local ternary patterns approach. The reducer is responsible for collecting the data and write it to the output file. The MapReduce job splits the input hipi image

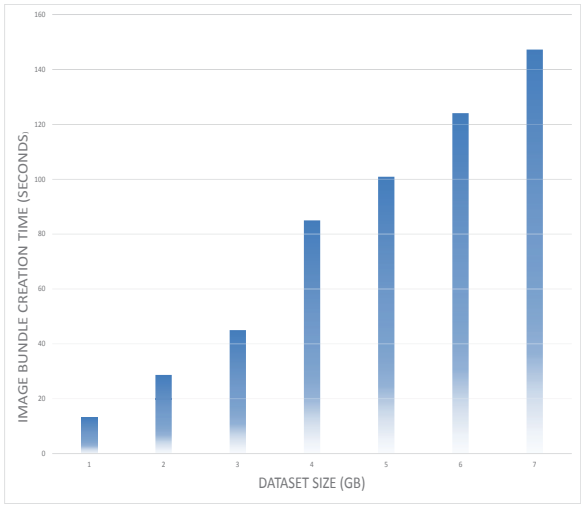

Figure 12: HIPI Image Bundle Creation Time

bundle into various data chunks. These data chunks then become input to the map tasks. The reducer gathers the processed data from the map tasks which are then stored in the database.

5.3.1 *Creating Hipi Image Bundle from Recorded Video Streams:*

The recorded video streams are H.264 encoded for faster transmission and efficient use of storage space. The video streams are first fetched from the video storage and are decoded to extract video frames from the input videos. The video decoding is performed by using the FFMPEG library. These decoded video frames are stored as PNGs. Each video stream is recorded at 25 frames per second. There are 3000 (=120*25) video frames for a video stream of 120 seconds length. The numbers of decoded video frames is dependent upon the length of a video stream being analyzed.
The individual frames are not suitable for further processing on the compute nodes. This is because of the fact that the MapReduce framework is developed to process large scale data and processing a small file will decrease the overall performance. These small files also necessitate lots of disk seeks and hopping time from node to node. Therefore, the decoded frames are first bundled using the HIPI Image Bundle [28] and are then processed by the map-reduce framework.

5.3.2 *Hipi Image Bundle Creation Time with Varying Data Sets:*

In order to measure the Hipi Image Bundle creation time, we have varied the datasets from 1GB to 7GB. These data sets with varying sizes assisted in evaluating different features of the system. These data sets are converted into one hipi image bundle before passing on to the map reduce framework. The hipi image bundle creation time varied between 13.18 seconds to 147.26 seconds for 1GB to 7GB datasets respectively. Figure 12 shows the time required to transform various input datasets into an image bundle. It can be noted that the time needed to create an image bundle increases with the increasing size of the data set.

We have also measured the time required to transfer the decoded video streams to the cloud nodes. After transferring

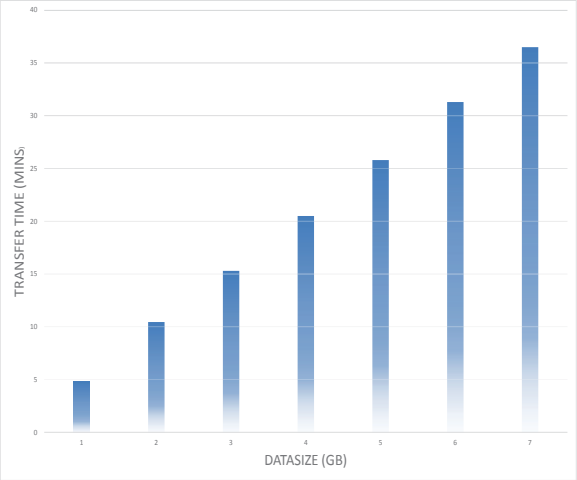

Figure 13: Date Transfer Time to Cloud Nodes

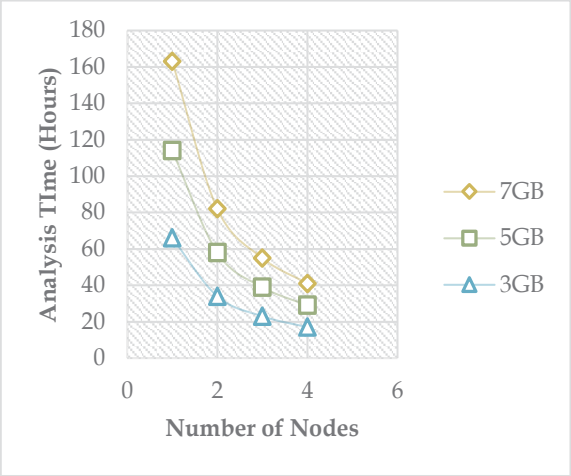

Figure 14: Analysis Time on Cloud Nodes with Varying Datasets

the video streams to the cloud nodes, the hipi image bundle is created. The transfer time of a dataset to the cloud nodes is dependent on the network bandwidth and the cloud data storage block size. For the datasets in our experiments (1GB - 7GB), the data transfer time varied from 4.8 minutes to 36.5 minutes. Figure 13 shows the data transfer time for varying data sizes.

5.3.3 Analysing Video Streams on Cloud Nodes:

The scalability of the proposed system is evaluated by executing it on the cloud nodes. We have analyzed datasets ranging from 1GB to 7GB on the cloud nodes. The time required to analyze datasets on the cloud nodes is measured to evaluate the performance of the system. It is observed that with an increase in the dataset size, an increasing trend is observed in the execution time (Figure 14). The number of spawned map tasks has a heavy impact on the performance of the overall system. The numbers of map tasks are dependent on the amount of data and the corresponding input split size. The maximum number of map tasks on a compute node depends on the input data set, the cloud data storage block size and the available hardware specifications of the node. The results show that performance of the system increases by distributing and parallelizing the work across multiple compute nodes. In this particular setup the default input split size of 128 MB is used.

As the number of nodes decreases, the amount of analysis tasks on each node increases which ultimately reduces the performance of the overall system. The degradation in the performance occurs because each task waits for longer period of time to get scheduled on the compute nodes. The analysis task has a minimum execution time and it is not possible to minimize the time beyond a certain limit. This is because of the inter process communication and the data read and write operation for the cloud storage. The proposed approach on a single node takes 163 hours to analyze a dataset of 7GB, while the same dataset is analyzed in 41 hours with 4 nodes. So a decreasing trend is observed in the execution time with an increasing number of nodes.

6. CONCLUSION AND FUTURE WORK

A cloud based blur and illumination invariant object classification system is presented and evaluated in this paper. The proposed system overcomes the challenges of blur and illumination by employing a stack based hierarchy of local ternary patterns generated from the amplitude spectrum of intrinsic mode functions of each input frame. The accuracy of the proposed system is demonstrated by experimentations on video data as well as publically available image datasets. The system proved to outperform state of the art approaches under controlled conditions. In order to demonstrate the scalability of the system, it is deployed on a cloud based infrastructure. The system proved to scale with increasing volumes of data on an increasing number of nodes. The larger volumes of data require more analysis time. However, the analysis time would decrease with the addition of more nodes to the cloud.

In future, we aim to extend the approach to cope with rotation and translation challenges. We would also experiment the system on a much bigger dataset with large number of cloud nodes. The integration of proposed system with deep learning approaches will also be the part of our future work.

7. REFERENCES

[1] Ojansivu, Ville, and Janne Heikkilä. “Blur insensitive texture classification using local phase quantization.” International conference on image and signal processing. Springer Berlin Heidelberg, 2008.

[2] Samad, Saleha, and Anam Haq. “Orientation Invariant Object Recognitions Using Geometric Moments Invariants and Color Histograms.” International Journal of Computer and Electrical Engineering, 2015.

[3] Joost Van de Weijer, Cordelia Schmid. “Blur robust and color constant image description”. International Conference on Image Processing (ICIP ’06), 2006.

[4] Wang, Jing-Wein, Ngoc Tuyen Le, Jiann-Shu Lee, and Chou-Chen Wang. “Color face image enhancement using adaptive singular value decomposition in fourier domain for face recognition.” Pattern Recognition, 2016.

[5] Zhang, Dehai, Da Ding, Jin Li, and Qing Liu. “Pca based extracting feature using fast fourier transform for facial expression recognition.” In Transactions on Engineering Technologies, Springer, 2015.

[6] Sagar GV, Barker SY, Raja KB, Babu KS, Venugopal KR. “Convolution based Face Recognition using DWT and feature vector compression.” In Third International Conference on Image Information Processing (ICIIP) 2015.
[7] Saini, Nirmala, and Aloka Sinha. “Face and palmprint multimodal biometric systems using GaborâĂŞWigner transform as feature extraction.” Pattern Analysis and Applications, 2015.
[8] Mandic, Danilo P., Naveed ur Rehman, Zhaohua Wu, and Norden E. Huang. “Empirical mode decomposition-based time-frequency analysis of multivariate signals: the power of adaptive data analysis.” IEEE Signal Processing Magazine, 2013.
[9] Chen WK, Lee JC, Han WY, Shih CK, Chang KC. “Iris recognition based on bidimensional empirical mode decomposition and fractal dimension.” Information Sciences. 2013.
[10] M.J. Swain and D.H. Ballard, “Color indexing,.” International journal of computer vision, 1991.
[11] B.V. Funt and G.D. Finlayson, “Color constant color indexing.” IEEE Trans. on Pattern Analysis and Machine Intelligence, 1995.
[12] J. Flusser and T. Suk. “Degraded image analysis: An invariant approach.” IEEE Trans. Pattern Anal. Machine Intell., 1998.
[13] Crosswhite, Nate, Jeffrey Byrne, Omkar M. Parkhi, Chris Stauffer, Qiong Cao, and Andrew Zisserman. “Template adaptation for face verification and identification.”, 2016.
[14] Aggarwal, Ashutosh, and Chandan Singh. “Zernike Moments-Based Gurumukhi Character Recognition.” Applied Artificial Intelligence, 2016.
[15] V. Ojansivu and J. HeikkilÂĺa. “Object recognition using frequency domain blur invariant features.” In Proc. Scandinavian Conference on Image Analysis (SCIAâĂŹ07), 2007.
[16] Ville Ojansivu and Janne Heikkila. “A Method for Blur and Similarity Transform Invariant Object Recognition.” In Proceedings of the 14th International Conference on Image Analysis and Processing (ICIAP ’07). IEEE Computer Society, 2007.
[17] Ville Ojansivu , Janne Heikkilä, “Blur Insensitive Texture Classification Using Local Phase Quantization”, Proceedings of the 3rd international conference on Image and Signal Processing, 2008.
[18] Grasso M, Chatterton S, Pennacchi P, Colosimo BM. “A data-driven method to enhance vibration signal decomposition for rolling bearing fault analysis.” Mechanical Systems and Signal Processing, 2016.
[19] Wadhwa N, Rubinstein M, Durand F, Freeman WT. “Riesz pyramids for fast phase-based video magnification.” IEEE International Conference on Computational Photography (ICCP), 2014.
[20] Freitas, Pedro Garcia, Welington YL Akamine, and Mylene CQ Farias. “No-reference image quality assessment based on statistics of Local Ternary Pattern.” Eighth International Conference on Quality of Multimedia Experience (QoMEX), 2016.
[21] http://hadoop.apache.org/
[22] http://opencv.org/
[23] http://hipi.cs.virginia.edu/
[24] https://www.bioid.com/About/BioID-Face-Database
[25] http://vision.ucsd.edu/content/yale-face-database
[26] Ashiq Anjum, Tariq Abdullah, Muhammad Tariq, Yousaf Baltaci and Nick Antonopoulos, “Video Stream Analysis in Clouds: An Object Detection and Classification Framework for High Performance Video Analytics”, IEEE Transactions on Cloud Computing, 2016.
[27] Tariq Abdullah, Ashiq Anjum, M. Fahim Tariq, Yusuf Baltaci and Nikos Antonopoulos, “Traffic Monitoring Using Video Analytics in Clouds”, 7th International Conference on Utility and Cloud Computing, 2017.
[28] Muhammad Usman Yaseen, Muhammad Sarim Zafar, Ashiq Anjum and Richard Hill, “High Performance Video Processing in Cloud Data Centres”, IEEE Symposium on Service-Oriented System Engineering (SOSE) 2016.