A Novel Image Retrieval Technique using Automatic and Interactive Segmentation

Asjad Amin and Muhammad Qureshi

Department of Telecommunication Engineering, The Islamia University of Bahawalpur, Pakistan

Abstract: *In this paper, we present a new region-based image retrieval technique based on robust image segmentation. Traditional content-based image retrieval deals with the global description of a query image. We combine the state-of-the-art segmentation algorithms with the traditional approach to narrow the area of interest to a specific region within a query image. In case of automatic segmentation, the algorithm divides a query image automatically and computes Zernike moments for each region. For interactive segmentation, our proposed scheme takes as input a query image and some information regarding the region of interest. The proposed scheme then works by computing the Geodesic-based segmentation of the query image. The segmented image is our region of interest which is then used for computing the Zernike moments. The Euclidean distance is then used to retrieve different relevant images. The experimental results clearly show that the proposed scheme works efficiently and produces excellent results.*

Keywords: *CBIR, information retrieval, image segmentation, multimedia image retrieval.*

1. Introduction

Digital images form a major part of today's internet world. Content Based Image Retrieval (CBIR) algorithms match a query image to a large number of images from various databases based on the contents of the query image. The efficiency of CBIR algorithms depends on the accuracy of the retrieved images. A number of CBIR approaches have been proposed, some of which involve user interaction. Most of such algorithms ask the user to provide its input in the form of relevance feedback while in some of the CBIR algorithms; the user gives additional information about the content of the query image.

In classical CBIR algorithms, a user is required to give one query image. The algorithm processes the query image and compares it to the database images to retrieve similar images. The CBIR algorithms that consider global view of the query image work well if the query image contains low visual color and texture information. However, such algorithms fail to achieve high accuracy rate when the query image contains rich color and texture information. A new concept of localized content based image retrieval or Region Based Image Retrieval (RBIR) was therefore introduced.

In localized CBIR, the user provides some additional information about content of query image. This additional information is about region of query image. The RBIR algorithms refine query image based on regional information provided by the user. In other types of localized CBIR, the query image is automatically divided into different regions. The content of refined query image is used to extract features which are then used to retrieve relevant images. Feature extraction methods in RBIR algorithms are based on color or texture of refined and selected regions. Most common features used in RBIR are Moment, Zernike moments descriptors, Pseudo Zernike moments, Generic Fourier descriptors, Grid Descriptors, Scale invariant feature transform, among others.

In this paper, we propose a localized image retrieval technique that is based on automatic and interactive segmentation. Automatic segmentation is implemented using the dynamic region merging approach proposed in [16] and further discussed in [6]. Interactive segmentation is implemented using Geodesic distance [1, 2]. For feature extraction, we have used traditional Zernike transform. The rest of the paper is structured as follows. We present the related work in section 2 and proposed image retrieval method in section 3. Then we discuss the experiment results and present comparisons with the state-of-the-art in section 4 followed by conclusion in section 5.

2. Related Work

Content based image retrieval has emerged as a challenging problem with a lot of potential for people working in the field of image processing and therefore, it has drawn a lot of attention. Although a number of algorithms are present in literature for image searching but the efficiency of these algorithms is low as compared to search algorithms for text. Traditionally, images are indexed using some

keywords. The keywords are assigned on the basis of image contents. Although the process of search using keywords is very fast but the accuracy of retrieved images significantly depends upon the accuracy of the keywords. Another problem with textual annotations based technique is that assigning keywords to an image is very complex, slow and an expensive task [19].

A more efficient approach for image retrieval is to exploit the visual contents of an image for indexing and searching. Content based image retrieval deals with algorithms that use visual information and perceptual content of an image. Color, shape, and texture are the most important visual properties of an image. CBIR uses features based on the content of an image for searching.

Shape is an important geometrical feature of an image that remains unaffected due to translation, rotation, and scaling. A lot of CBIR techniques use shape descriptors to estimate visual content of images. The shape descriptors based techniques can be classified into two main categories i.e., contour-based, and region-based techniques. The contour-based techniques use the information extracted from the edges of an image. The contour-based techniques use mainly Fourier descriptors [25] and contour point distribution matching [19]. However, the boundary or edge information alone is not enough for image searching.

Unlike contour-based techniques, region-based techniques consider the complete image content for the process of image searching. Region-based techniques are robust to noise and other similar variations. Region-based techniques include moment based methods or distribution based descriptors in which histograms represent various characteristics of the region. Moment-based methods include Zernike Moments Descriptors (ZMD) [10], Generic Fourier Descriptors (GFD) [24], etc., Zernike moments are widely used in image retrieval. However, they involve a radial kernel function that contains a factorial term and therefore leads to high computational cost especially for higher orders.

The distribution-based methods include Scale Invariant Feature Transform (SIFT) [13], histograms of oriented gradients [5], gradient location and orientation histogram [15], etc., The drawback with distribution-based descriptors is that they are computationally intensive and produce high dimensional feature vectors.

The CBIR algorithms that consider global view of the query image work well if the query image contains less texture information. To overcome this, the localized content based image retrieval or RBIR was introduced. In localized CBIR, the user is only interested in a specific region of the image. To localize an area of interest, the user can either mark the desired area or in some cases it is determined by the CBIR algorithm. The efficiency of such algorithms depends on the accurate estimation of the area of interest.

Many CBIR algorithms divide the image into different blocks [14]. The size of each block is predefined. Features such as Zernike moments are computed for each block. Some CBIR algorithms extract salient points [9, 11] such as SIFT points. Features are only computed for the chosen points. Such algorithms rely on the image points that give maximum efficiency. However, these algorithms do not really take into account the area of interest.

Some localized CBIR techniques [17, 22] use more generalized partitioning schemes such as segmentation to accurately classify the area of interest. Features are extracted only from the desired segmented region. Segmentation techniques are classified in two general categories i.e., automatic segmentation and interactive segmentation. In automatic segmentation [16], input image is divided into different meaningful regions according to some predefined criteria. This scheme may not be suitable for localized CBIR as it does not segment the desired area. In interactive segmentation [2], user is required to provide some information regarding the desired area. The user can mark some pixels from the desired area and some pixels from the background. On the basis of this information, the algorithm extracts the desired area which is the region of interest in this case. The remaining image is excluded. Other CBIR techniques are discussed in [3, 4, 7, 8, 18, 20]. The CBIR techniques using deep learning [12, 21] are also gaining popularity. However, these techniques require adequate training which is not suitable for many retrieval applications.

There are very few papers that use segmentation to improve the retrieval accuracy of query image. This has inspired us to choose this topic. Here, we propose an image retrieval method in which we use robust segmentation to refine the area of interest within a query image. We use both automatic and interactive robust image segmentation for image retrieval.

3. Proposed Methodology

We propose a localized image retrieval scheme based on automatic and interactive segmentation. Automatic segmentation is applied to all the images of the database. The Zernike features are then computed for each region and stored. For a given query image, the user can either chose automatic or interactive segmentation. In the case of automatic segmentation, the same procedure as mentioned before will be followed. In case of interactive segmentation, Geodesic distance is computed with the help of user interaction, where the user provides some information related to the desired and undesired regions in the query image. After getting the segmented image, the algorithm computes the Zernike moments. The Euclidean distance is then used to retrieve the matched images. Figure 1 shows the flow chart of our proposed algorithm.
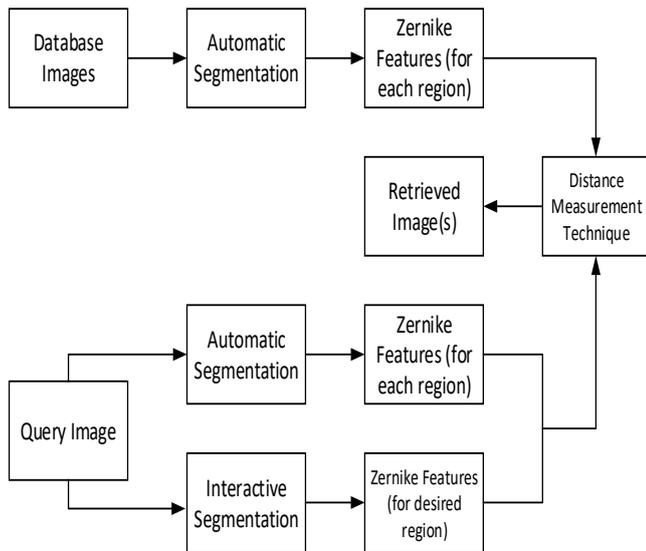
Figure 1. The work flow of the proposed methodology.

3.1. Automatic Segmentation based on Region Merging

For automatic segmentation, we have used The Dynamic Region Merging (DRM) [16]. In dynamic region merging, we divide the image into a set of over segmented regions. To improve computational efficiency, we chose small regions instead of single pixel. We combine these small regions using a pre-defined predicate.

The predicate used in dynamic region merging is based on the dissimilarity between pixels along the boundary of two regions. For this, an image is represented as an undirected graph in which each node represents a region. Each edge, E, has a corresponding weight which indicates the dissimilarity of the two regions. The merging predicate $P(R_i, R_j)$ on regions R_i and R_j is defined as merge R_i and R_j if they are the most similar neighbours in each other's neighbourhood. Figure 2 shows an example of over segmented region and its realization using region adjacency graph.
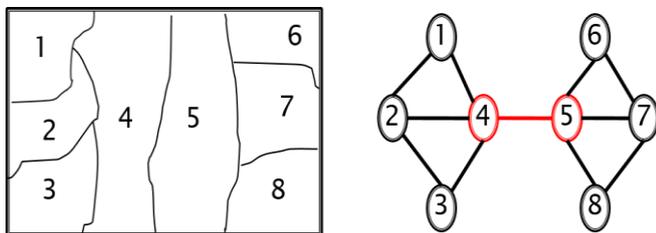

Figure 2. Over segmented region (Left) and its equivalent adjacency graph (Right).

The DRM is implemented using the following steps:

1. Divide the input image into n over-segmented regions.
2. Compute the predicate P for each region, S_i, with its neighbours.
3. Create new segmented regions S_{i+1} by merging the neighbouring regions with P=true.
4. Repeat 2 & 3 until $S_{i+1} = S_i$.

5. Return the final segmented image.

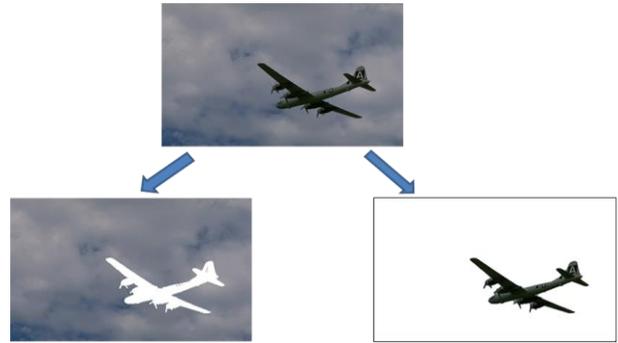Figure 3 shows two segmented regions of the aeroplane image using automatic segmentation.

Figure 3. Segmentation of Aero plane image using automatic segmentation.

3.2. Interactive Segmentation based on Geodesic Distance

In interactive segmentation [6], the geodesic distance is computed using the initial information provided by user. To compute the geodesic distance, we implement kernel density estimation using Gauss transform, based on some information provided by the user about the desired region. Next, we compute the minimum distance from all the pixels to the desired pixels selected by the user. The Geodesic distance is then computed using kernel density estimation and shortest path. The geodesic distance for each pixel is the smallest integral of a weight function over all paths from the foreground and background points [2].

3.2.1. Kernel Density Estimation (KDE)

The KDE uses some kernel to estimate the density function of color images. Computing density function of color images is computationally expensive and time-consuming. The Gauss transform, as discussed in [23], is one of the efficient ways to compute KDE. The algorithm presented in [23] can be used with color images and provides a computationally efficient solution. In Figure 4, we show an example of KDE for water bear image. Next, we compute conditional KDE using foreground and background pixels.
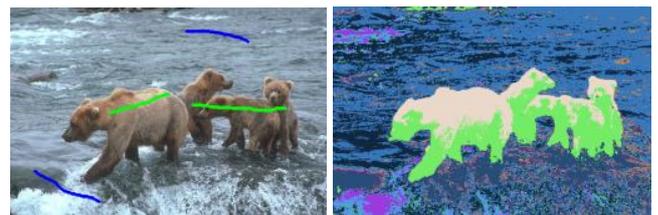

Figure 4. Original image (left), kernel density of image using gauss transform (right).

3.2.2. Shortest Path for Each Pixel from Foreground and Background Pixels

Finally, we compute the shortest path, using the Dijkstra's algorithm, from all pixels to the foreground and background pixels marked by the user. For each pixel in foreground and background set, the Dijkstra's algorithm finds the lowest cost path from that pixel to all the other pixels. The following steps are performed to compute shortest path.

1. Initialize: Set the Costs (C) of all the pixels=∞
2. Pick the starting point from the set of pixels marked by the user.
3. For each pixel, do
   - C(neighbor)=min(C(pixel)+C(path_pixel_to_neighbor), C (neighbor))
   - Add neighbor to the REMAINING set
4. Select a new pixel from REMAINING set with minimum cost.
5. Go back to step 3 and repeat for all the pixels.

Figure 5 shows the original image and final segmented image with our desired region.

Figure 5. Original image (left), final segmented image (right).

3.3. Zernike Moments

The Zernike moments were introduced in 1990 by Tague. They are considered to be robust features for image analysis and pattern recognition problems. Computing the Zernike moments for an image is a three step process. In the first step, radial polynomials are computed. In the second step, Zernike basis functions are computed. In the last step, Zernike moments are computed by projecting the image on to the basis functions [10]. The real valued 1-D radial polynomial $R_{n,m}$ is given as:

$$R_{nm}(r) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s! \times (\frac{n+|m|}{2} - s)! \times (\frac{n-|m|}{2} - s)!} r^{n-2s} \quad (1)$$

Where n is the order of the radial polynomial. High order Zernike moments introduce high complexity and they are also sensitive to noise. Therefore, the selection of n and m should be made in a way that the resulting Zernike moments are of low order. Table 1 gives the value of m and n that we have used for this paper.

Table 1. Description of zernike moments used in the proposed technique.

N	M	Total Moments
3	1, 3	
4	0, 2, 4	
5	1, 3, 5	12
6	0, 2, 4, 6	

4. Experimental Results

4.1. Dataset Description

For our experiments, we have used the Corel draw database for content based image retrieval. The database consists of 10,800 images divided into 80 different groups. Figure 6 shows some images from the Corel database.

Figure 6. Images from corel database.

4.2. Experimental Results

To illustrate the performance of the proposed algorithm, we have chosen four different pictures from the Corel dataset as shown in Figure 7. For each image, we search through the Corel database to retrieve relevant images using the traditional Zernike transform method, proposed method using automatic segmentation, and proposed method using interactive segmentation. Figure 8 shows the top eight retrieved images for the aeroplane image, shown in Figure 7-a, using Zernike transform. We see that some of the images are not relevant and the performance of traditional CBIR algorithm using Zernike moments is not satisfactory. Figure 9 shows the automatic segmentation of the aeroplane image. The image is segmented into three main regions. Now, we use this result to search through the Corel database. The top eight retrieved images, using automatic segmentation, are shown in Figure 10. It can be seen that the proposed method provides much better results with better accuracy.

a) Aeroplane.   b) Butterfly.   c) Goat.   d) Light house.

Figure 7. Images used for experiments.

Figure 8. Image retrieved for Figure 7 (a) using traditional zernike moments.
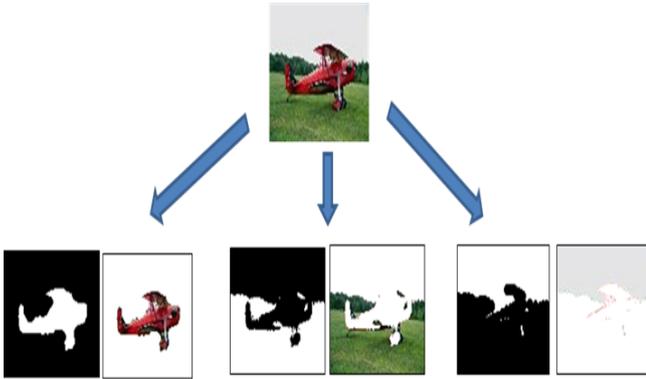

Figure 9. Automatic segmentation of Figure 7 (a) using dynamic region merging.

Figure 10. Image retrieved for Figure 7 (a) using proposed method based on automatic segmentation.

Figures 11 and 12 show the search results using traditional Zernike transform based method and proposed method using automatic segmentation for goat image shown in Figure 7-c. Figure 13 shows the results if we use interactive segmentation. Here, we search only for the goat, ignoring the rest of image.

Figure 11. Image retrieved for Figure 7 (c) using zernike moments.

Figure 12. Image retrieved for Figure 7 (c) using proposed method based on automatic segmentation.

Figure 13. Image retrieved for Figure 7 (c) using proposed method based on interactive segmentation.

Figures 14 and 15 show another example from the Corel database. In Figure 14, we search for query image using proposed algorithm based on automatic segmentation. In Figure 15, we use interactive segmentation and search only for the desired area i.e., light house.

Figure 14. Image retrieved for figure 7 (d) using proposed method based on automatic segmentation.

Figure 15. Image retrieved for Figure 7 (d) using proposed method based on interactive segmentation.

We measure the performance of the Image retrieval systems in terms of its precision. Precision measures the ability of the system to retrieve only the relevant images. Precision is given as: Precision=Number of relevant images retrieved/Total number of images retrieved.

We show in Table 2 the performance of the proposed system and the other state-of-the art image retrieval systems for four different categories chosen from Corel Database. We see that the proposed method using interactive segmentation outperforms all the other techniques. The cost we pay here is the extra input that is required from the user for interactive segmentation. The performance of the proposed method using automatic segmentation is also very

encouraging. Except for the Goat category, this scheme outshines all the other techniques.

Table 2. Average precision of retrieval by different methods for aeroplane, goat, lighthouse and butterfly category.

Category Name	Proposed Method using Automatic Segmentation	Proposed Method using Interactive Segmentation	Traditional Zernike based method [10]	RBIR using watershed algorithm [4]	Wavelet based color histogram WBCH
Aeroplane	0.79	0.84	0.49	0.74	0.71
Goat	0.85	0.91	0.53	0.80	0.87
Lighthouse	0.93	0.96	0.82	0.90	0.92
Butterfly	0.88	0.92	0.46	0.86	0.76

5. Conclusions

We discussed in this paper a new approach for image retrieval using a combination of robust image segmentation and the traditional Zernike moments. Contrary to traditional CBIR approaches, we propose here to analyze only special regions of interest obtained from segmentation. Our experimental results showed that the proposed algorithm achieves excellent retrieval accuracy. The search results were tested using the comprehensive Corel database. The proposed image retrieval algorithm is applicable to all types of images whereas four images have been presented here as cases studies. The experimental analysis clearly suggests that the proposed algorithm produces better results than existing techniques. The algorithm is fast and can be used as an excellent processing tool in online image retrieval applications.

References

[1] Amin A. and Deriche M., "Robust Image Segmentation Based on Convex Active Contours and the Chan Vese Model," *in Proceedings of IEEE Global Conference on Signal and Information Processing*, Atlanta, pp. 1044-1048, 2014.

[2] Bai X. and Sapiro G., "Geodesic Matting: A Framework for Fast Interactive Image and Video Segmentation and Matting," *International Journal of Computer Vision*, vol. 82, no. 2, pp. 113-132, 2009.

[3] Barnard K. and Shirahatti N., "Method for Comparing Content Based Image Retrieval Methods," *in Proceedings of Internet Imaging IV*, Santa Clara, pp. 1-9, 2003.

[4] Chiang C., Hung Y., Yang H. and Lee G., "Region-Based Image Retrieval Using Color-Size Features of Watershed Regions," *Journal of Visual Communication and Image Representation*, vol. 20, no. 3, pp. 167-177, 2009.

[5] Dalal N. and Triggs B., "Histograms of Oriented Gradients for Human Detection," *in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, pp. 886-893, 2005.

[6] Deriche M., Amin A., and Qureshi M., "Color Image Segmentation by Combining The Convex Active Contour and the Chan Vese Model," *Pattern Analysis and Applications*, vol. 22, no. 2, pp. 1-15, 2017.

[7] Howarth P. and Ruger S., "Evaluation of Texture Features for Content-Based Image Retrieval," *in Proceedings of International Conference on Image and Video Retrieval*, Dublin, pp. 326-334, 2004.

[8] Karuppusamy J. and Marappan K., "Efficient Color and Texture Feature Extraction Technique for Content Based Image Retrieval System," *The International Arab Journal of Information Technology*, vol. 13, no. 6, pp. 784-790, 2016.

[9] Ke Y. and Sukthankar R., "PCA-SIFT: A more Distinctive Representation for Local Image Descriptors," *in Proceedings of IEEE conference on Computer Vision and Pattern Recognition*, Washington, pp. 1-8, 2004.

[10] Kim W. and Kim Y., "A Region Based Shape Descriptor Using Zernike Moments," *Signal Processing: Image Communication*, vol. 16, no. 1-2, pp. 95-102, 2000.

[11] Ledwich L. and Williams S., "Reduced Sift Features for Image Retrieval and Indoor Localisation," *in Proceedings of Australasian Conference on Robotics and Automation*, Canberra, pp. 1-4, 2004.

[12] Lin K., Yang H., Hsiao J., and Chen C., "Deep Learning of Binary Hash Codes for Fast Image Retrieval," *in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Boston, pp. 27-35, 2015.

[13] Lowe D., "Distinctive Image Features from Scale-Invariant Key Points," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.

[14] Maron O. and Ratan A., "Multiple-Instance Learning for Natural Scene Classification," *in Proceedings of the 15th International Conference on Machine Learning*, San Francisco, pp. 341-349, 1998.

[15] Mikolajczyk K. and Schmid C., "A Performance Evaluation of Local Descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615-1630, 2005.

[16] Peng B., Zhang L., and Zhang D., "Automatic Image Segmentation by Dynamic Region Merging," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3592-3605, 2011.

[17] Rahmani R., Goldman S., Zhang H., Krettek J., and Fritts J., "Localized Content-Based Image Retrieval," *in Proceedings of the 7th ACM SIGMM International Workshop on Multimedia*

Information Retrieval, Hilton, pp. 227-236, 2005.

[18] Shanmugam B., Rathinavel R., Perumal T., and Subbaiyan S., "An Efficient Perceptual of CBIR System using MIL-SVM Classification and SURF Feature Extraction," *The International Arab Journal of Information Technology*, vol. 14, no. 4, pp. 428-435, 2017.

[19] Shu X. and Wu X., "A Novel Contour Descriptor for 2D Shape Matching and its Application to Image Retrieval," *Image and Vision Computing*, vol. 29, no. 4, pp. 286-294, 2011.

[20] Smeulders A. W., Worring M., Santini S., Gupta A., and Jain R., "Content-Based Image Retrieval at The End of The Early Years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349-1380, 2000.

[21] Wan J., Wang D., Hoi S., Wu P., Zhu J., Zhang Y., and Li J., "Deep learning for Content-Based Image Retrieval: A Comprehensive Study," *in Proceedings of the 22nd ACM International Conference on Multimedia*, Orlando, pp. 157-166, 2014.

[22] Wang J., Li J., and Wiederhold G., "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 9, pp. 947-963, 2001.

[23] Yang C., Duraiswami R., Gumerov N., and Davis L., "Improved Fast Gauss Transform and Efficient Kernel Density Estimation," *in Proceedings of the 9th IEEE International Conference on Computer Vision*, Nice, pp. 664-671, 2003.

[24] Zhang D. and Lu G., "Shape-based Image Retrieval Using Generic Fourier Descriptor," *Signal Processing: Image Communication*, vol. 17, no. 10, pp. 825-848, 2002.

[25] Zhang D. and Lu G., "A Comparative Study of Curvature Scale Space and Fourier Descriptors for Shape Based Image Retrieval," *Journal of Visual Communication and Image Representation*, vol. 14, no. 1, pp. 39-57, 2003.

Asjad Amin received the B.S. degree in telecommunication engineering from NU-FAST, Islamabad, Pakistan, M.S. degree in electrical engineering from University of Engineering & Technology, Taxila, Pakistan and Ph.D. in electrical engineering from King Fahd University of Petroleum & Minerals, Saudi Arabia in 2017. Since 2008, he is with the department of telecommunication engineering at The Islamia University of Bahawalpur, currently serving as Assistant Professor. His research interests include image and video processing, seismic imaging and modeling, machine learning, and image segmentation.

Muhammad Qureshi is currently working as Assistant Professor in Telecommunication Engineering Department, University College of Engineering & Technology, The Islamia University of Bahawalpur. He received his B.Sc. degree in electrical engineering from UET Lahore in 2000 and M.Sc. in telecommunication engineering from NWFP-UET Peshawar in 2008. He completed his Ph.D. in Electrical Engineering from KFUPM, Saudi Arabia in 2017 His research interests include image & video processing, image compression, image forensics, and image quality assessment. He has published 18 refereed international journals and conference papers. Muhammad Ali Qureshi is a senior member of IEEE and reviewer of many renowned international journals with good impact factors in the related discipline.