Handling Uncertainty under Spatial Feature Extraction through Probabilistic Shape Model (PSM)

Ahmed, Samsuddin[1]; Rahman, Md. Mahbubur[2]

[1, 2] Bangladesh University of Business and Technology, Department of Computer Science and Engineering, Rupnagar Residential Area, Mirpur 2, Dhaka-1216, Bangladesh

Abstract: ***The Extraction of Spatial Features from remotely sensed data and the use of this information as input into further decision making systems such as geographical information systems (GIS) has received considerable attention over the few decades. The successful use of GIS as a decision support tool can only be achieved, if it becomes possible to attach a quality label to the output of each spatial analysis operation. Thus the accuracy of Spatial Feature Extraction gained more attention as geographic features can hardly formulated in a certain pattern due to intra-class variation and inter-class similarity. Besides these Spatial Feature Extraction further include positional uncertainty, attribute uncertainty, topological uncertainty, inaccuracy, imprecision/inexactitude, inconsistency, incompleteness, repetition, vagueness, noisy, omittance, misinterpretation, misclassification, abnormalities and knowledge uncertainty. To control and reduce uncertainty in an acceptable degree, a Probabilistic shape model is described for Extracting Spatial Features from multi-spectral image. The advantages of this, as opposed to the conventional approaches, are greater accuracy and efficiency, and the results are in a more desirable form for most purposes.***

Keywords: Spatial Feature, GIS, Computer Vision, Feature Extraction.

1. Introduction

An image contains enormous information. For clear representation of information embedded in image it can be decomposed into several features. The term ‘feature’ refers to scene objects (e.g. Road, River, Building, Tree, and Human) with similar characteristics (whether they are spectral, spatial or otherwise). Therefore, spatial feature extraction aims at detecting and locating objects in images without input from an operator on its initial position. As a reference detecting a building in images automatically has two tasks, i.e., recognition of a building and determination of its position. Recognizing a building in an image is much more difficult than determining its position as it requires not only the information which can be derived from the image, but also a priori knowledge about the properties of a building and its relationships with other features in the image and other related knowledge such as knowledge on the imaging system. Due to the complexity of images and existence of image noise and disturbances, the information derived from the image is always incomplete and ambiguous. These sorts of uncertainty make the recognition process more complex. This work mainly gives emphasis on uncertainty under spatial feature extraction.

The paper is organized in following order. Section A addresses the problem statement, basic concepts, the aims & objectives of the work along with the design considered and scope of the work. Section B reviewed the concepts related with this work. Section C presents a novel methodology for detecting and localizing objects representing spatial feature in cluttered image. Section D highlights on the dataset. Section E describes the evaluation of results of this work by concluding a comparison with other methods. Section F concludes the project work with some future research direction.

2. Literature Review

For last couple of years Spatial Feature Extraction by computer has been an active area of research. For much of time, the Spatial Feature Extraction approach has been dominated by the discovery of geometric and analytic representations of objects that can be used to predict the appearance of an object under any viewpoint and under any conditions of illumination and partial occlusion. There are a number of reasons why geometry has played a central role in spatial object detection framework.

1. Invariance to viewpoint: Geometric object descriptions allow the projected shape of an object to be accurately predicted under perspective projection.
2. Invariance to illumination: Recognizing geometric descriptions from images can be achieved using edge detection and geometric boundary segmentation. Such descriptions are reasonably invariant to illumination variations.
3. Well developed theory: Geometry has been under active investigation by mathematicians for thousands of years. The geometric framework has achieved a high degree of maturity and effective algorithms exist for analyzing and manipulating geometric structures.
4. Concentration of detection: Major concentration of objection detection was on the man-made objects. A large fraction of manufactured objects are designed using computer-aided design (CAD) models and therefore are naturally described by primitive geometric elements, such as planes and spheres. More complex shapes are also represented with simple geometric descriptions, such as a triangular mesh or polynomial patches.

The Blocks World was first established theoretical framework for cognitive tasks that simplifies objects to polyhedral shapes on a uniform background. But the framework made too many assumptions in recognition strategies that could not be expected to hold in real world scenes. However the blocks world avoided numerous difficulties such as:

1. Curved surfaces and boundaries;
2. Articulated and moving objects;
3. Occlusion by unknown shapes;
4. Complex background and 3-d texture such as foliage;
5. Specular or mutually illuminating surfaces;
6. Multiple light sources and remote shadowing;
7. Transparent or translucent surfaces.

These difficulties generated a period of passim PSM concerning the completeness and stability of bottom-up segmentation processes. Recent approaches to detect spatial feature fall into one of three major categories. The first category consists of systems that are model-based, i.e., a model is defined for the object of interest and the system attempts to match this model to different parts of the image in order to find a fit [13]. Most object detection systems use motion information, explicit models, and a static camera, assume a single person in the image, or implement tracking rather than pure detection. The second type are image invariance methods which base a matching on a set of image pattern relationships (e.g., brightness levels) that, ideally, uniquely determine the objects being searched for.[13].

The final set of object detection systems are characterized by the example-based learning algorithms. These systems learn the salient features of a class from sets of labeled positive and negative examples. Example-based techniques have also been successfully used in the area of computer vision also, including object recognition [11]. Papageorgiou et al. have successfully employed example-based learning techniques to detect people in complex static scenes without assuming any a priori scene structure or using any motion information. Haar wavelets [5] are used to represent the images and Support Vector Machine (SVM) classifiers [8] are used to classify the patterns. However, Papageorgiou's system's ability to detect partially occluded objects whose parts have little contrast with the background is limited. Some problems associated with Papageorgiou’s detection system may be addressed by taking a component-based approach to detecting objects. A component-based object detection system is one that searches for an object by looking for its identifying components rather than the whole object. These systems have two common features: They all have component detectors that identify candidate components in an image and they all have a means to integrate these components and determine if together they define an object.

It is worth mentioning that a component-based object detection system for spatial feature extraction is harder to realize than one for a single object because the geometry of different objects varies dramatically. This means that not only is there greater intra-class variation concerning the configuration of an object, but also that it is more difficult to detect object parts in the first place since their appearance can change significantly in place to place.

Many current object detection methods deal with this problem by performing an exhaustive search over all possible object positions and scales. This exhaustive search imposes severe constraints, both on the detector’s computational complexity and on its discriminance, since a large number of potential false positives need to be excluded. An opposite approach is to let the search be guided by image structures that give cues about the object scale. In such a system, an initial interest point detector tries to find structures whose extent can be reliably estimated under scale changes. These structures are then combined to derive a comparatively small number of hypotheses for object locations and scales. Only those hypotheses that pass an initial plausibility test need to be examined in detail. In recent years a range of scale-invariant interest point detectors have become available which can be used for this purpose.

In our approach, we combine several of the above ideas. Our system uses a large number of automatically selected parts, based on the output of an interest point operator, and combines them flexibly in a star topology. Robustness to scale changes is achieved by employing scale-invariant interest points and explicitly incorporating the scale dimension in the hypothesis search procedure. The whole approach is optimized for efficient learning and accurate detection from small training sets. The probabilistic shape model is used to remove uncertainty under spatial feature extraction.

3. Methodology

Though current approaches of object detection reached to a level to identify a large number of previously seen and known objects, it is still not well-understood to recognize unseen-before objects of a given category and assigning the correct category label. Obviously, this task is more difficult, since it requires a method to cope with large within class variations of object colors, textures, and shapes, while retaining at the same time enough specificity to avoid mis-classifications. In order to learn the appearance variability of an object category, a vocabulary of local appearances is built up that are characteristic for (a particular viewpoint of) its member objects. This is done by extracting local features around interest points and grouping them with an agglomerative clustering scheme. Based on this vocabulary, a Probabilistic shape model (PSM) is learnt that specifies where on the object the vocabulary entries may occur. The advantages of this approach are its greater flexibility and the smaller number of training examples it needs to see in order to learn possible object shapes. For example, when learning to categorize articulated objects such as building or trees, this method does not need to see every possible articulation in the training set. This idea is similar in spirit to approaches that represent novel objects by a combination of class prototypes [9], or of familiar object views [11]. However, the main difference of this approach is that here the combination does not occur between entire exemplar objects, but through the use of local image features, which again allows a greater flexibility. Directly connected to the recognition procedure, a probabilistic formulation is derived for the top-down segmentation problem, which integrates learned knowledge of the recognized category with the supporting information in the image. The resulting procedure yields a pixel-wise figure-ground segmentation as a result and extension of recognition. In addition, it delivers a per-pixel confidence estimate specifying how much this segmentation can be trusted. The automatically computed top-down segmentation is then in turn used to improve recognition. First, it allows only aggregating evidence over the object region and discarding influences from the background. Second, the information from where in the image a hypothesis draws its support makes it possible to resolve ambiguities between overlapping hypotheses. Minimum Description Length (MDL) principle is used to formalize this idea. The resulting procedure constitutes a novel mechanism that allows analyzing image containing spatial features. The whole approach is formulated in a scale-invariant manner, making it applicable in real-world situations where the object scale is often unknown.

3.1 Vocabulary Building

A vocabulary is built based on local appearances that are characteristic for a certain viewpoint of an object category by sampling local features that repeatedly occur on a set of training images of this category. The basic idea is inspired by the work of [12], [6], Building a vocabulary includes three consecutive steps. These are;

1. Interest Point Detector
2. Interest Point Descriptor
3. Grouping similar Feature

Step one and two cojugately known as local feature extraction.

In the following, steps of vocabulary generation procedure are described.

3.1.1 Interest Point Detector

A scale-invariant interest point detector is applied to obtain a set of informative regions for each image. By extracting features only from those regions, the amount of data to be processed is reduced, while the interest point detector's preference for certain structures assures that "similar" regions are sampled on different objects. Several different interest point detectors are available for this purpose. In this project, Harris [14], Harris-Laplace [1], Hessian-Laplace [1], and Difference-of-Gaussian (DoG) [7] detectors are used.

3.1.2 Interest Point Descriptor

The extracted image regions are represented by a local descriptor. Several descriptor choices are available for this step. In this project, Grey-value Patches [2], SIFT [7], and Local Shape Context [4], [3] descriptors are used. In order to develop the different stages of this approach regions detected and descripted are simply referred by the term feature.

3.1.3 Grouping Similar Features

Visually similar features are grouped to create a vocabulary of prototypical local appearances. In order to keep the representation as simple as possible, representation of all features in a cluster is done by their mean which is the cluster center. A necessary condition for this is that the cluster center is a meaningful representative for the whole cluster. In that respect, it becomes evident that the goal of the grouping stage must not be to obtain the smallest possible number of clusters, but to ensure that the resulting clusters are visually compact and contain the same kind of structure. This is an important consideration to bear in mind when choosing the clustering method. Therefore in this project agglomerative clustering schemes are used. This approach automatically determines the number of clusters by successively merging features until a cut-off threshold t on the cluster compactness is reached.

3.2 Learning the Shape Model

Let C be the learned appearance codebook, as described in the previous section. The next step is to learn the spatial probability distribution PC. For this, we perform a second iteration over all training images and match the codebook entries to the images. Here, we activate not only the best-matching codebook entry, but all entries whose similarity is above t, the cut-off threshold already used during agglomerative clustering. For every codebook entry, we store all positions it was activated in, relative to the object center. By this step, we model the uncertainty in the codebook generation process. If a codebook is "perfect" in the sense that each feature can be uniquely assigned to exactly one cluster, then the result is equivalent to a nearest-neighbor matching strategy. However, it is unrealistic to expect such clean data in practical applications. We therefore keep each possible assignment, but weight it with the probability that this assignment is correct. It is easy to see that for similarity scores smaller than t, the probability that this patch could have been assigned to the cluster during the codebook generation process is zero; therefore those matches do not need to consider. The stored occurrence locations, on the other hand, reflect the spatial distribution of a codebook entry over the object area in a non-parametric form.

3.3 Recognition of Spatial Features

Given a test image, an interest point detector is again applied and features are extracted around the selected locations. The extracted features are then matched to the codebook to activate codebook entries using the mechanism as described above. From the set of all those matches, we collect consistent configurations by performing a Generalized Hough Transform. Each activated entry casts votes for possible positions of the object center according to the learned spatial distribution Pc. Consistent hypotheses are then searched as local maxima in the voting space. When pursuing such an approach, it is important to avoid quantization artifacts. Once a hypothesis has been selected, all patches that contributed to it are collected, thereby visualizing what the system reacts to. As a result, a representation of the object including a certain border area is got. This representation can optionally be further refined by sampling more local features. The back projected response will later serve as the basis for computing a category-specific segmentation, as described later.

3.3.1Probabilistic Hough Voting

Let f be an evidence for an extracted image feature observed at location l. By matching it to the codebook, a set of valid interpretations C_i with probabilities p (C_i|f,) can get. If a codebook cluster matches, it casts votes for different object positions. That is, for every C_i, we can obtain votes for several object categories/viewpoints on and positions x, according to the learned spatial distribution P $(o_n,x|C_i)$.

3.3.2 Top down segmentation

The local maxima / hypothesis support already provides a rough indication where the object is in the image. As we have expressed the unknown image content in terms of a learned codebook; we know more about the semantic interpretation of the matched patches for the target object. In the following, it is shown how this information can be used to infer a pixel-wise figure-ground segmentation of the object. In order to learn this top-down segmentation, this approach requires a reference figure-ground segmentation for the training images. While this additional information might not always be available, we will demonstrate that it can be used to improve recognition performance significantly. As a consequence of non-parametric representation for PC, the resulting algorithm is very simple and can be efficiently computed.

3.4 Hypothesis Verification

The previous section has shown that we can obtain probabilistic top-down segmentation from each hypothesis and thus split its support into figure and ground pixels. The basic idea of this verification stage is to only aggregate evidence over the figure portion of the image that is over pixels that are hypothesized to belong to the object, and discard misleading information from the background. The motivation for this is that correct hypotheses will lead to consistent segmentations, since they are backed by an existing object in the image. False positives from random background clutter, on the other hand, will often result in inconsistent segmentations and thus in lower figure probabilities. At the same time, this idea allows to compensate for a systematic bias in the initial voting scheme. The probabilistic votes are constructed on the principle that each feature has the same weight. This leads to a competitive advantage for hypotheses that contain more matched features simply because their area was more densely sampled by the interest point detector. Normalizing a hypothesis's score by the number of contributing features, on the other hand, would not produce the desired results, because the corresponding image patches can overlap and may also contain background structure. By accumulating evidence now over the figure pixels, the verification stage removes this over-counting bias. Using this principle, each pixel has the same potential influence, regardless of how many sampled patches it is contained in. Finally, this strategy makes it possible to resolve ambiguities from overlapping hypotheses in a principled manner. When applying the recognition procedure to real-world test images, a large number of the initial false positives are due to secondary hypotheses which overlap part of the object. This is a common problem in object detection that is particularly prominent in scenes containing multiple objects. Generating such secondary hypotheses is a desired property of a recognition algorithm, since it allows the method to cope with partial occlusions. However, if enough support is present in the image, the secondary detections should be suppressed in favor of other hypotheses that better explain the image. Usually, this problem is solved by introducing a bounding box criterion and rejecting weaker hypotheses based on their overlap. However, such an approach may lead to missed detections. Again, using the top-down segmentation this system can improve misdetection and exactly quantify how much support the overlapping region contains for each hypothesis. In particular, this permits us to detect secondary hypotheses, which draw all their support from areas that are already better explained by other hypotheses, and distinguish them from true overlapping the principle of Minimal Description Length (MDL), which combines all of those motivations.
The complete algorithm is described below.

Algorithm: The Probabilistic Shape Model Algorithm

Input: Test Image
Output: Spatial Feature

1. Initialize the set of probabilistic votes as null.
2. Apply the interest point detector to the test image.
3. for all interest regions $l_k= (i,l_x, l_y, l_s)$ with descriptors f_k repeat step 4 to step 12
4. Initialize the set of matches M as null
5. for all codebook entries C_i repeat step 6
6. if similarity between feature descriptor f_k and vocabulary entry C_i is greater than or equal to threshold value then record a match by $M \leftarrow M \cup (i,l_x, l_y, l_s)$
7. for all matching codebook entries C_i^* set the match weight by $p(C_i^*|f_k) \leftarrow 1/|M|$
8. for all matches $(i, l_x,l_y, l_s) \in M$ reapeat step 9
9. for all occurrences occ ∈ Occ[i] of codebook entry C_i repeat step 10 to12
10. Set the vote location by $x \leftarrow (lx- occ_x*(l_s/occ_s), l_y - occ_y*(l_s/occ_s), l_s/occ_s)$
11. Set the occurrence weight by $p(o_n,x|C_i, l) \leftarrow 1/|Occ[i]|$
12. Cast a vote (x, w, occ,l) for position x with weight w by $w \leftarrow p(on, x|C_i, l)p(C_i|f_k)$ and $V \leftarrow V \cup (x, w, occ, l)$
13. Given hypothesis h and supporting votes V_h for all supporting votes $(x, w, occ, l) \in V_h$, let img_{mask} be the segmentation mask corresponding to occ and sz be the size at which the interest region was sampled. Repeat step 14 to step 15
14. Rescale img_{mask} to sz by computing $u_0 \leftarrow (x - 0.5sz)$ and $v_0 \leftarrow (y - 0.5sz)$
15. for all $u \in [0, sz - 1]$ repeat step 16
16. for all $v \in [0, sz - 1]$ repeat step 17 and step 18
17. $img_{pf\ ig}(u - u_0, v- v_0)$+=$w \bullet img_{mask}(u, v)$
18. $img_{pgnd}(u - u_0, v- v_0)$+=$w \bullet (1 - img_{mask}(u, v))$

4. Result

In this chapter Google Earth image are analyzed to demonstrate the feasibility of the proposed system. This approach is evaluated on the building, road and water-land categories in mentioned dataset. We chose these categories since the dataset consists of a large number of images for these categories and many other methods have reported results on these categories. The proposed method gives better result than others.

Throughout the entire work, it started with gathering images for feature extraction. This step is followed by the multispectral feature extraction that the project is pursuing for implementing spatial feature extraction. The feature extraction on Google earth image of ICT Building of University of Chittagong is given below.

Figure 1: Test image of CU.

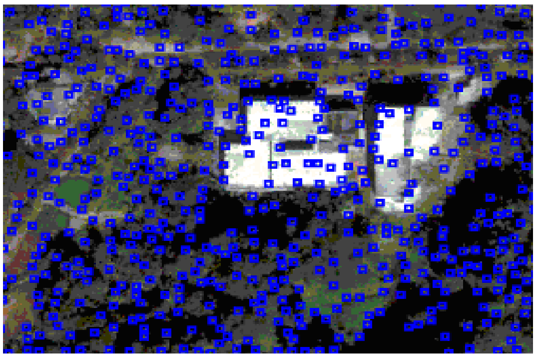

Figure 2: Interest point detected

Figure 2 shows the extracted features for an example image (in this case using Harris interest points). As can be seen from this example, the sampled information provides a dense cover of the object, leaving out only uniform regions. This process is repeated for all training images, and the extracted features are collected

By step by step computation the proposed method recognizes the building and segmented it from background. The percentage of recognition is summarized in the following table 5.1.

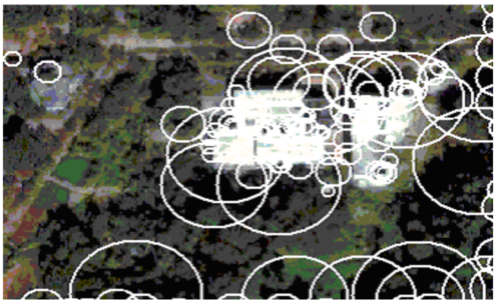

Figure 3: Probabilistic Hough Voting

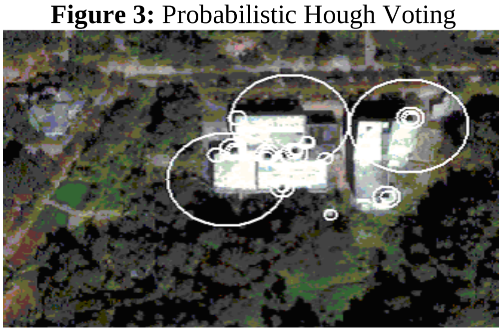

Figure 4: Hypothesis Verification

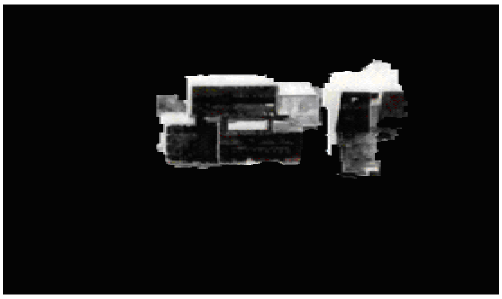

Figure 5: Top down segmentation

Table 1: Result

Name of Spatial Feature / Object	*No of Training Image*	*Recognition Accuracy*
Building	10	92%
Road	12	90.15%
Tree	5	95.06%
Human	3	98.25%

4.3 Discussion

In the above table we observed that objects with specified part and shape are best recognized. As the shape of river and road is not so different these can't be separated yet. If their valency is considered it may be possible to recognize. May be cloud model or random forest model can be used to do it. Overall the result is satisfactory as it outperforms than.. .This method needs very small training set to detect spatial features.

5. Conclusion

Spatial feature extraction is a very significant and fundamental concept in Computer Vision and GIS. It also stands on a considerable and elementary idea of image processing. The huge amount of information embedded in images need to be apparently represented. For this purpose, the features are originated. Then the objects i,e spatial features are detected. This project aimed for extracting the spatial features / objects for providing a fundamental abstraction for modeling the structure of images representing various raster images. The core of this project is an established probabilistic shape model that is carried out for spatial feature extraction or object detection. As the work continues, the project tries to implement every part of the procedure so as to reduce the time and space complexity and to establish its effectiveness and efficiency. The project also made some comparative study between various methods in computer to reach at the correct decision in case of choosing a process.

The level best attempt to develop this project lead the entire aim to reach at some satisfactory position though it entails some limitations also. These limitations are considered as the future work. This can be considered as the extension of the current project.

6. Future Work

The project will attempt to recover the uncertainty in detecting spatial features occurs from the variations of shape, color , texture etc in same type of objects and similarities of the factors mentioned for different class of objects. We have crucial intra class variations as well as similarities among different classes. So it will be the major concentration of the project to develop a system that will detect objects classes efficiently and accurately. Moreover, more generalization needs to be done for spatial feature extraction in case of gathering the satellite images from any distance.

References

[1] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. "A comparison of af?ne region detectors." IJCV, 65(1/2):43–72, 2005.

[2] S. Agarwal, A. Atwan, and D. Roth. "Learning to detect objects in images via a sparse, part-based representation". Trans. PAMI, 26(11):1475–1490,

[3] K. Mikolajczyk and C. Schmid. "A performance evaluation of local descriptors." Trans. PAMI, 27(10), 2005.

[4] A. Garg, S. Agarwal, and T. Huang. "Fusion of global and local information for object detection." In ICPR'02, 2002.

[5] D. Magee and R. Boyle. "Detecting lameness using 're-sampling condensation' and 'multi-stream cyclic hidden markov models'. Image and Vision Computing, 20(8):581–594, 2002.

[6] M. Weber, M. Welling, and P. Perona. "Towards automatic discovery of object categories." In CVPR'00, 2000.

[7] D. Lowe. "Distinctive image features from scale-invariant key points." IJCV, 60(2):91–110, 2004.

[8] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio, "Pedestrian Detection Using Wavelet Templates" Proc. Computer Vision and Pattern Recognition, pp. 193-199, June 1997.

[9] M. Jones and T. Poggio. "Model-based matching by linear combinations of prototypes," MIT AI Memo 1583, MIT, 1996.

[10] S. Ullman. "Three-dimensional object recognition based on the combination of views." Cognition, 67(1):21–44, 1998.

[11] H. Murase and S. Nayar, "Visual Learning and Recognition of 3D Objects from Appearance" Int'l J. Computer Vision, vol. 14, no. 1, pp. 5-24, 1995.

[12] M.C. Burl, M. Weber, and P. Perona. "A probabilistic approach to object recognition using local photometry and global geometry." In ECCV'98,

[13] P. Sinha, "Object Recognition via Image Invariants: A Case Study," Investigative Ophthalmology and Visual Science, vol. 35, pp. 1735-1740, May 1994.

[14] C. Harris and M. Stephens. "A combined corner and edge detector". In Alvey Vision Conference, pages 147–151, 1988.

Author Profile

Samsuddin Ahmed has been lecturing in CSE since mid of 2010. He was graduated in Computer Science and Engineering from University of Chittagong with highest CGPA till date. His hobbies include thinking about underlying mathematical formulations in natural phenomena. His research interests embrace data and image mining, Semantic Web, Business Intelligence, Spatial Feature Extraction etc. He is now serving one of the top most private Universities in Bangladesh named Bangladesh University of Business and Technology (BUBT).

Md.Mahbubur Rahman, received his B.Sc. Engineering degree in computer science and engineering from Patuakhali Science and Technology University in 2011. He is now working toward his M.Sc. Engineering degree in computer science and engineering at Bangladesh University of Engineering and Technology (BUET), Bangladesh and working as a Lecturer in computer science and engineering at one of the top most private Universities in Bangladesh named Bangladesh University of Business and Technology (BUBT). His research interests are Data mining, Search Engine optimization, Biometric system, Network Security.