TERRIAN IDENTIFICATION USING CO-CLUSTERED MODEL OF THE SWARM INTELLEGENCE & SEGMENTATION TECHNIQUE

Ritesh Srivastava[1] Shivani Agarwal[2] Ankit Goel[3] Vipul Gupta[4]

[1]Department of Computer Science ,Galgotia College, Greater Noida.
ritesh21july@gmail.com[1]
[2,3,4]Department of Computer Science, IMS Engineering College, Ghaziabad.
kasishivani@gmail.com[2] ankitgoel100@gmail.com[3]
vipulgp1987@gmail.com[4]

ABSTRACT

A digital image is nothing more than data -- numbers indicating variations of red, green, and blue at a particular location on a grid of pixels. Clustering is the process of assigning data objects into a set of disjoint groups called clusters so that objects in each cluster are more similar to each other than objects from different clusters. Clustering techniques are applied in many application areas such as pattern recognition, data mining, machine learning, etc. Clustering algorithms can be broadly classified as Hard, Fuzzy, Possibility, and Probabilistic .K-means is one of the most popular hard clustering algorithms which partitions data objects into k clusters where the number of clusters, k, is decided in advance according to application purposes. This model is inappropriate for real data sets in which there are no definite boundaries between the clusters. After the fuzzy theory introduced by Lotfi Zadeh, the researchers put the fuzzy theory into clustering. Fuzzy algorithms can assign data object partially to multiple clusters. The degree of membership in the fuzzy clusters depends on the closeness of the data object to the cluster centers. The most popular fuzzy clustering algorithm is fuzzy c-means (FCM) which introduced by Bezdek in 1974 and now it is widely used. Fuzzy c-means clustering is an effective algorithm, but the random selection in center points makes iterative process falling into the local optimal solution easily. For solving this problem, recently evolutionary algorithms such as genetic algorithm (GA), simulated annealing (SA), ant colony optimization (ACO) , and particle swarm optimization (PSO) have been successfully applied.

KEYWORDS:

Image Segmentation, medical imaging, super pixels, Particle Swarm Optimization, FCM, Swarm Intellegence.

Natarajan Meghanathan, et al. (Eds): ITCS, SIP, JSE-2012, CS & IT 04, pp. 239–249, 2012.

DOI : 10.5121/csit.2012.2122

1. INTRODUCTION

A digital image is nothing more than data -- numbers indicating variations of red, green, and blue at a particular location on a grid of pixels. Clustering is the process of assigning data objects into a set of disjoint groups called clusters so that objects in each cluster are more similar to each other than objects from different clusters. Clustering techniques are applied in many application areas such as pattern recognition , data mining , machine learning , etc. Clustering algorithms can be broadly classified as Hard, Fuzzy, Possibilistic, and Probabilistic .K-means is one of the most popular hard clustering algorithms which partitions data objects into k clusters where the number of clusters, k, is decided in advance according to application purposes. This model is inappropriate for real data sets in which there are no definite boundaries between the clusters.After the fuzzy theory introduced by Lotfi Zadeh, the researchers put the fuzzy theory into clustering. Fuzzy algorithms can assign data object partially to multiple clusters. The degree of membership in the fuzzy clusters depends on the closeness of the data object to the cluster centers. The most popular fuzzy clustering algorithm is fuzzy c-means (FCM) which introduced by Bezdek in 1974 and now it is widely used. Fuzzy c-means clustering is an effective algorithm, but the random selection in center points makes iterative process falling into the local optimal solution easily. For solving this problem, recently evolutionary algorithms such as genetic algorithm (GA), simulated annealing (SA), ant colony optimization (ACO) , and particle swarm optimization (PSO) have been successfully applied.

2. ANT COLONY OPTIMIZATION

In the natural world, ants (initially) wander randomly, and upon finding food return to their colony while laying down pheromone trails. If other ants find such a path, they are likely not to keep travelling at random, but to instead follow the trail, returning and reinforcing it if they eventually find food (see Ant communication).Over time, however, the pheromone trail starts to evaporate, thus reducing its attractive strength. The more time it takes for an ant to travel down the path and back again, the more time the pheromones have to evaporate. A short path, by comparison, gets marched over more frequently, and thus the pheromone density becomes higher on shorter paths than longer ones. Pheromone evaporation also has the advantage of avoiding the convergence to a locally optimal solution. If there were no evaporation at all, the paths chosen by the first ants would tend to be excessively attractive to the following ones. In that case, the exploration of the solution space would be constrained.

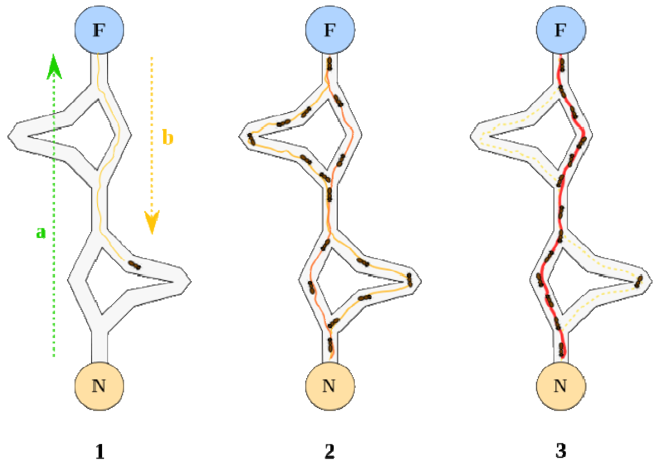


Fig 1: Ants Behaviour

3. PARTICLE SWARM OPTIMIZATON

In computer science, **particle swarm optimization** (**PSO**) is a computational method that optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. PSO optimizes a problem by having a population of candidate solutions, here dubbed particles, and moving these particles around in the search-space according to simple mathematical formulae over the particle's position and velocity. Each particle's movement is influenced by its local best known position and it's also guided toward the best known positions in the search-space, which are updated as better positions are found by other particles. This is expected to move the swarm toward the best solutions.

PSO is originally attributed to Kennedy, Eberhart and Shi and was first intended for simulating social behaviour, as a stylized representation of the movement of organisms in a bird flock or fish school. The algorithm was simplified and it was observed to be performing optimization. The book by Kennedy and Eberhart[4] describes many philosophical aspects of PSO and swarm intelligence. An extensive survey of PSO applications is made by Poli. PSO is a metaheuristic as it makes few or no assumptions about the problem being optimized and can search very large spaces of candidate solutions. However, metaheuristics such as PSO do not guarantee an optimal solution is ever found. More specifically, PSO does not use the gradient of the problem being optimized, which means PSO does not require that the optimization problem be differentiable as is required by classic optimization methods such as gradient descent and quasi-newton methods. PSO can therefore also be used on optimization problems that are partially irregular, noisy, change over time, etc.

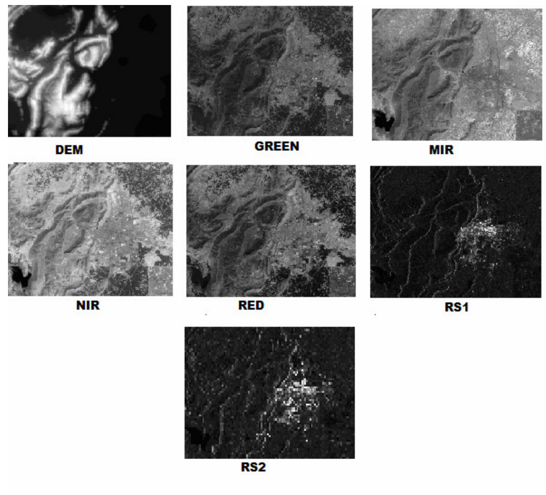


Fig 2: Seven Band Images of Alwar

3.1 PARAMETER VALUE SELECTION

For the initial version of the PSO, the values for c_1, c_2, and v_{max} have to be selected. This selection has an impact on the convergence speed and the ability of the algorithm to find the optimum, but different values may be better for different problems. Much work has been done to select a combination of values that works well in a wide range of problems. For the constricted version of PSO, the following restrictions are proposed [5]:

$$\chi = \frac{2}{\left|2-\varphi-\sqrt{\varphi^2-4\cdot\varphi}\right|} \quad (1)$$

$$\varphi = c_1 + c_2 > 4.0 \quad (2)$$

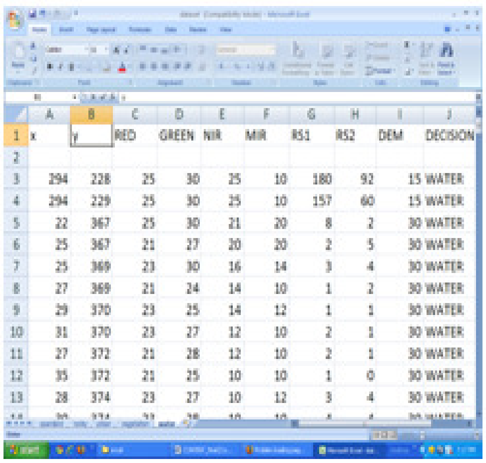
Regarding the inertia weight, it determines how the previous velocity of the particle influences the velocity in the next iteration:

If w=0, the velocity of the particle is only determined by the p_i and p_g positions; this means that the particle may change its velocity instantly if it is moving far from the best positions in its knowledge. Thus, low inertia weights favor exploitation (local search).If w is high, the rate at which the particle may change its velocity is lower (it has an "inertia" that makes it follow its original path) even when better fitness values are known. Thus, high inertia weights favor exploration (global search).

c_1	c_2	w_{init}	w_{final}
2.0	2.0	1.0	1.0
2.0	2.0	0.9	0.4
1.4962	1.4962	0.7968	0.7968

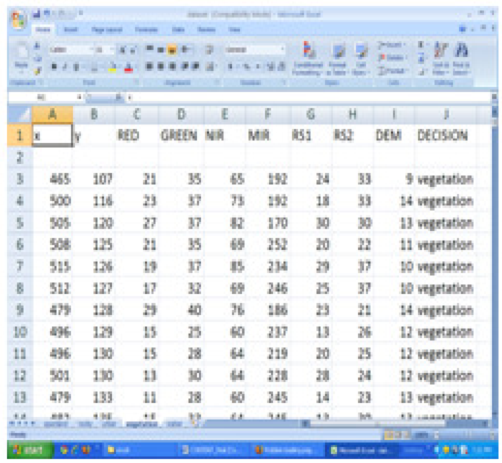
3.2 INPUT DATA SETS

3.2.1 Data Set of water region



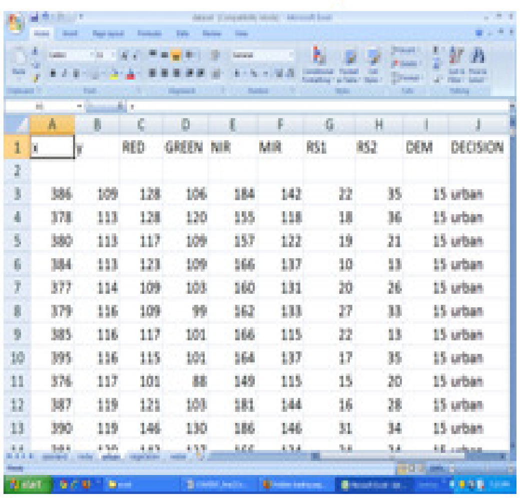
x	y	RED	GREEN	NIR	MIR	RS1	RS2	DEM	DECISION
294	228	25	30	25	10	180	92	15	WATER
294	229	25	30	25	10	157	60	15	WATER
22	367	25	30	21	20	8	2	30	WATER
25	367	21	27	20	20	2	5	30	WATER
25	369	23	30	16	14	3	4	30	WATER
27	369	21	24	14	10	1	2	30	WATER
29	370	23	25	14	12	1	1	30	WATER
31	370	23	27	12	10	2	1	30	WATER
27	372	21	28	12	10	2	1	30	WATER
35	372	21	25	10	10	1	0	30	WATER
28	374	23	27	10	12	3	4	30	WATER

3.2.2 Data set for Vegitation Region



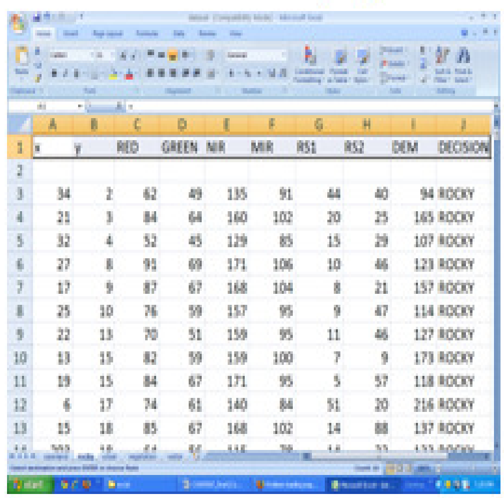
x	y	RED	GREEN	NIR	MIR	RS1	RS2	DEM	DECISION
465	107	21	35	65	192	24	33	9	vegetation
500	116	23	37	73	192	18	33	14	vegetation
505	120	27	37	82	170	30	30	13	vegetation
508	125	21	35	69	252	20	22	11	vegetation
515	126	19	37	85	234	29	37	10	vegetation
512	127	17	32	69	246	25	37	10	vegetation
479	128	29	40	76	186	23	21	14	vegetation
496	129	15	25	60	237	13	26	12	vegetation
496	130	15	28	64	219	20	25	12	vegetation
501	130	13	30	64	228	28	24	12	vegetation
479	133	11	28	60	245	14	23	13	vegetation

3.2.3 Data set for Urban Region



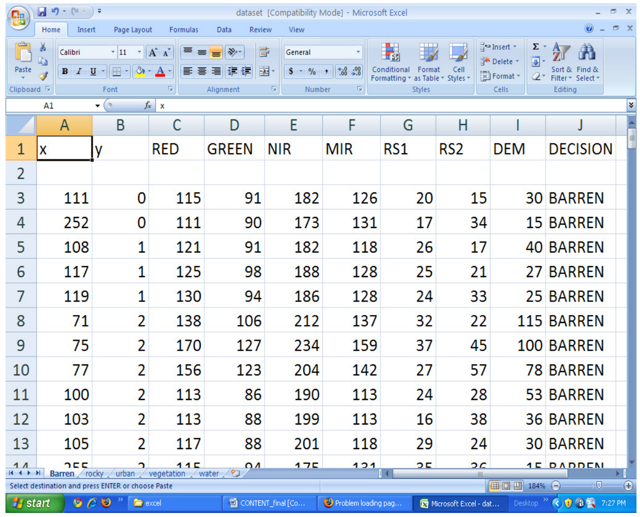
x	y	RED	GREEN	NIR	MIR	RS1	RS2	DEM	DECISION
386	109	128	106	184	142	22	35	15	urban
378	113	128	120	155	118	18	36	15	urban
380	113	117	109	157	122	19	21	15	urban
384	113	123	109	166	137	10	13	15	urban
377	114	109	103	160	131	20	26	15	urban
379	116	109	99	162	133	27	33	15	urban
385	116	117	101	166	115	22	13	15	urban
395	116	115	101	164	137	17	35	15	urban
376	117	101	88	149	115	15	20	15	urban
387	119	121	103	181	144	16	28	15	urban
390	119	146	130	186	146	31	34	15	urban

3.2.4 Data set for Rocky Region



x	y	RED	GREEN	NIR	MIR	RS1	RS2	DEM	DECISION
34	2	62	49	135	91	44	40	94	ROCKY
21	3	84	64	160	102	20	25	165	ROCKY
32	4	52	45	129	85	15	29	107	ROCKY
27	8	91	69	171	106	10	46	123	ROCKY
17	9	87	67	168	104	8	21	157	ROCKY
25	10	76	59	157	95	9	47	114	ROCKY
22	13	70	51	159	95	11	46	127	ROCKY
13	15	82	59	159	100	7	9	173	ROCKY
19	15	84	67	171	95	5	57	118	ROCKY
6	17	74	61	140	84	51	20	216	ROCKY
15	18	85	67	168	102	14	88	137	ROCKY

3.2.5 Data set for Barren Region



x	y	RED	GREEN	NIR	MIR	RS1	RS2	DEM	DECISION
111	0	115	91	182	126	20	15	30	BARREN
252	0	111	90	173	131	17	34	15	BARREN
108	1	121	91	182	118	26	17	40	BARREN
117	1	125	98	188	128	25	21	27	BARREN
119	1	130	94	186	128	24	33	25	BARREN
71	2	138	106	212	137	32	22	115	BARREN
75	2	170	127	234	159	37	45	100	BARREN
77	2	156	123	204	142	27	57	78	BARREN
100	2	113	86	190	113	24	28	53	BARREN
103	2	113	88	199	113	16	38	36	BARREN
105	2	117	88	201	118	29	24	30	BARREN

4. SEGMENTED IMAGE FILTERATION PROCESS

Image filtering basically deals with the filteration of all pixels in the image in different patches. Each patch contains a set of pixel corresponding to which filteration process takes place. Each segment passes through different stages of alteration of pixels, such as darkness, blue-ish, transparency etc. which then modifies the image over its different aspects. Set of characterstics features of image are dealt and filteration process separately and uniquely treats the patches, more over these patches or clusters of part of the whole image is sustained to alteration of features. Now lets us elaborate an example describing advancement in the field of cluster analysis and treatment of images.As Images are divided into separate parts that are homogeneous with respect to property of brightness, color or texture, etc. The large scale spatial image is a complex and complicated. For this, K-means clustering is proposed for segmentation in terms to the decision rule. Furthermore, special simplification case study of K-means clustering is presented to be an alternative in order to reduce the computational complexity.

5. MODIFICATION IN K-MEANS CLUSTERING ALGORITHM

The *k*-means algorithm assigns each point to the cluster whose center (also called centroid) is nearest. The center is the average of all the points in the cluster — that is, its coordinates are the arithmetic mean for each dimension separately over all the points in the cluster. After the clusters are marked dimensionally fit for identification assemble the clusters to a uniform database record. *Example:* The data set has three dimensions and the cluster has two points: $X = (x_1, x2, x3)$ and $Y = (y_1, y2, y3)$. Then the centroid Z becomes $Z = (z_1,z_2,z_3)$, where

$$z_1 = \frac{x_1 + y_1}{2}, \quad z_2 = \frac{x_2 + y_2}{2} \text{ and } z_3 = \frac{x_3 + y_3}{2}.$$

The algorithm steps are:

- Choose the number of clusters, *k*.
- Randomly generate *k* clusters and determine the cluster centers, or directly generate *k* random points as cluster centers.
- Assign each point to the nearest cluster center, where "nearest" is defined with respect to one of the distance measures discussed above.
- Now set out the sets of clusters and mark them done, those which are analysed and covered to some specific outcome.
- Repeat the procedure and specify all the identified clusters.
- Let us now combine the identified clusters via certain probability possibilities.
- Match up the corresponding clusters to make out certain already existing database.
- Repeat the two previous steps until some convergence criterion is met (usually that the assignment hasn't changed).

The main advantages of this algorithm are its simplicity and speed which allows it to run on large datasets. Its disadvantage is that it does not yield the same result with each run, since the resulting clusters depend on the initial random assignments (the k-means++ algorithm addresses this problem by seeking to choose better starting clusters). It minimizes intra-cluster variance, but does not ensure that the result has a global minimum of variance. Another disadvantage is the requirement for the concept of a mean to be definable which is not always the case. For such datasets the k-medoids variants is appropriate. An alternative, using a different criterion for which points are best assigned to which centre is k-medians clustering.

6. PROPOSED ALGORITHM

- For each particle $i = 1, ..., S$ do:
    - Initialize the particle's position with a uniformly distributed random vector: $\mathbf{x}_i \sim U(\mathbf{b}_{lo}, \mathbf{b}_{up})$, where $\mathbf{b}_{lo}$ and $\mathbf{b}_{up}$ are the lower and upper boundaries of the search-space.
    - Initialize the particle's best known position to its initial position: $\mathbf{p}_i \leftarrow \mathbf{x}_i$
    - If $(f(\mathbf{p}_i) < f(\mathbf{g}))$ update the swarm's best known position: $\mathbf{g} \leftarrow \mathbf{p}_i$
    - Initialize the particle's velocity: $\mathbf{v}_i \sim U(-|\mathbf{b}_{up}-\mathbf{b}_{lo}|, |\mathbf{b}_{up}-\mathbf{b}_{lo}|)$

- Until a termination criterion is met (e.g. number of iterations performed, or adequate fitness reached), repeat:
    - For each particle $i = 1, ..., S$ do:

- Pick random numbers: $r_p, r_g \sim U(0,1)$
- Update the particle's velocity by equation (1):

$$v_{id}^{t-1} = \chi(w \cdot v_{id}^{t} + c_1 \cdot \psi_1 \cdot (p_{id}^{t} - x_{id}^{t}) + c_2 \cdot \psi_2 \cdot (p_{gd}^{t} - x_{id}^{t}))$$

where

ω = inertia weight

χ =constriction factor

C_1 =coginitive parameter
C_2 =social parameter

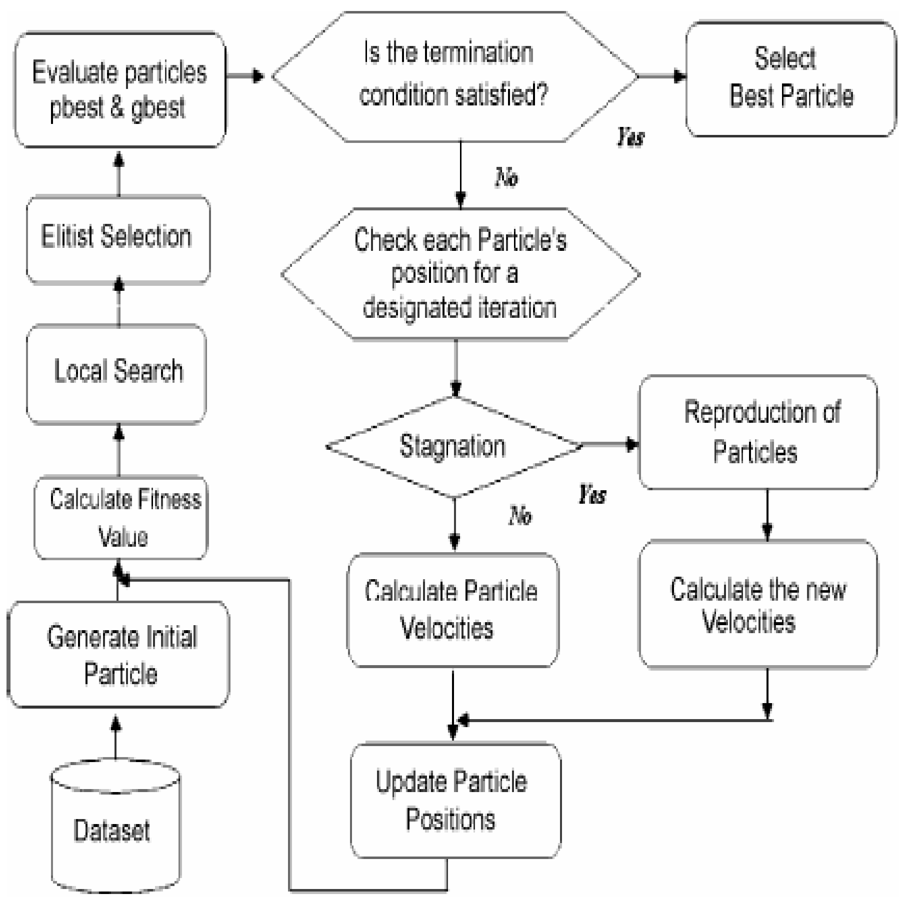
$$\chi = \frac{2K}{|2-\varphi-\sqrt{\varphi^2-4\varphi}\,|} \quad \text{for} \quad \varphi = C_1 + C_2 > 4$$

ELSE $\sqrt{K}$, K=1

- Update the particle's position: $\mathbf{x}_i \leftarrow \mathbf{x}_i + \mathbf{v}_i$
- If ($f(\mathbf{x}_i) < f(\mathbf{p}_i)$) do:
- Update the particle's best known position: $\mathbf{p}_i \leftarrow \mathbf{x}_i$
- If ($f(\mathbf{p}_i) < f(\mathbf{g})$) update the swarm's best known position: $\mathbf{g} \leftarrow \mathbf{p}_i$

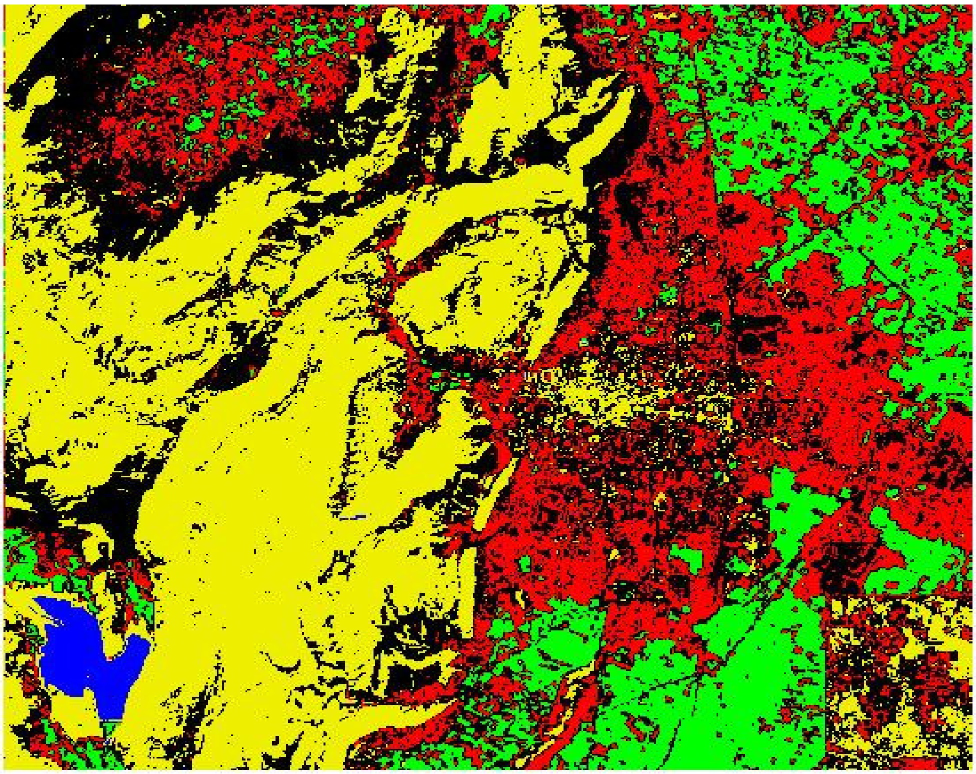
Now g holds the best found solution.

7. FLOW CHART



8. RESULT

8.1 OUTPUT IMAGE



8.2 ERROR MATRIX

	Veg	Urban	Rocky	Water	Barren	Total
Veg	127	1	1	0	0	129
Urban	1	92	0	0	39	129
Rocky	1	15	175	0	3	194
Water	0	0	0	70	0	70
Barren	21	81	26	0	128	256
Total	150	189	202	70	170	781

9. CONCLUSIONS

The complex problem of delineation of the Landuse/Landcover classes having similar spectral signatures depends extensively on expert's perception of the scene. Training sets, although a manifestation of the expert's knowledge is an indirect method. Most of the supervised classification techniques do not elicit the underlying knowledge contained in the training set.

It is observed that the Particle Swarm Optimization is a good technique and can be applied to classify the satellite image. The results are also good enough when compared to other complimentary techniques. Also it facilitates the explicit characterization of the underlying knowledge and provides a better insight into the classification mechanism.Further, the Ƙ -coefficient can also be used as a measure of elicited knowledge. This also provides us a mechanism to compare the knowledge content with other paradigms of supervised classification. Following inferences are arrived at:

i. The use of PSO, in eliciting expert's knowledge, a very complex task of resolving class conflict of spectrally similar classes which otherwise is usually avoided in the conventional methodologies, is innovative.
ii. Accuracy level of the classified image by this approach is comparable and quite acceptable, especially in time-scared scenario.
iii. It is very difficult even for an expert to explicitly explain his reasoning. The approach based on PSO offers a very transparent knowledge representation scheme.
iv. K-means clustering using criteria from defined distance metric, the satellite images (2m×2m/pixel) are processed and analyzed, while selected objects (e.g, rivers and trees) are successfully clustered.
v. Probability of more portable and clear visuable image is quite strong through k-means clustering algorithm.
vi. Clustering can be used to divide a digital image into distinct regions for border detection or object recognition.
vii. The *k*-means clustering algorithm is commonly used in computer vision as a form of image segmentation. The results of the segmentation are used to aid border detection and object recognition.

Finally, the elicited knowledge can be measured in terms of Ƙ-coefficient. This provides a very transparent and efficient mechanism for knowledge comparison with other black-box type of classification systems.

10. FUTURE SCOPE

The present technique is majorly a decision taking technique. The system performance can be increased by using better implementation methods and the heuristic functions. For better results it is advised to use better unsupervised classification methods. Also this algorithm already proves itself as a good competitor to its complementary techniques. Using this algorithm improvisation can be done in many applications. Also PSO performance itself can be increased by collaborating it with various other techniques.

REFERENCES

[1] Devis Tuia and Gustavo Camps-vallas, "Recent Advances In Remote Sensing Image Processing", IEEE Trans pp. 3705-3707, 2009.

[2] DEPSO: Hybrid Particle Swarm with Differential Evolution Operator. Wen-Jun Zhang, Xiao-Feng Xie* (Institute of Microelectronics,Tsinghua University,Beijing 100084, P. R. China)

[3] Convergence Behavior of the Fully Informed Particle Swarm Optimization Algorithm. (Institut de Recherches Interdisciplinaires et de Développements en Intelligence Artificielle)

[4] D. Bernstein, "Optimization r us," IEEE Control Systems Mag., vol. 26, pp. 6–7, 2006

[5] M. Dorigo, L. Gambardella, M. Middendorf, and T. Stutzle, Eds., "Special section on 'ant colony optimization'," IEEE Trans. Evol. Comput., vol. 6, no. 4, pp. 317–365, Aug. 2002.

[6] Reading the general concepts of the Particle Swarm Optimization technique by Mohamed A. El-Sharkawi ,Computational Intelligence Applications (CIA) Lab.University of Washington.

[7] M. Dorigo and T. Stutzle, Ant Colony Optimization (The MIT Press,2004).

[8] R. C. Eberhart and Y. Shi, "Particle swarm optimization: Developments,applications and resources," IEEE Int. Conf. Evolutionary Computation, vol. 1,pp. 81–86, 2001.

[9] Fuzzy Clustering Using Hybrid Fuzzy c-means and Fuzzy Particle Swarm Optimization,by Hesam Izakian, Ajith Abraham Machine Intelligence Research Labs MIR Labs,USA.and Václav Snášel Faculty of Electrical Engineering and Computer Science VSB-Technical University of Ostrava.

Authors:

Ritesh Srivastava is a Assistant Professor in Galgotia Engineering College, Greater Noida. He had done many projects on NLP & Image processing techniques, and have an interest in Research based on Soft Computing techniques.

Ms Shivani Agarwal is is a Assistant Professor in IMS Engineering College, Ghaziabad. She had done many projects on Soft Computing techniques, and have an interest in Research based on Soft Computing techniques and in Neural Network based Classification.

Ankit Goel is an B.tech final year student from IMS engineering college, having interest in research fiels based on image classifications and a member of Computer Society of India, chennai.

Vipul Gupta is an B.tech final year student from IMS engineering college, having interest in research fiels based on image classifications.