Task Scheduling of Firefly Algorithm based on Cloud Computing

Shuaili WANG*
Modern Education Technology Center, Hunan City University, Yiyang, Hunan, 413000, China

Abstract: Proposed a task scheduling in cloud computing based on intelligence firefly algorithm aimed at the disadvantages of cloud computing task scheduling. Firstly, on the basis of cloud model, used intelligence firefly algorithm with strong ability of global searching to find the better solution of cloud computing task scheduling then turned the better solution into the initial pheromone of improved firefly algorithm, and found out the cloud computing task scheduling and the algorithm's global optimal solution through improved firefly information communications and feedbacks. Finally, made comparison test of the three benchmark function on the basis of MATLAB, the results showed, compared with traditional intelligence firefly algorithms, the improved algorithm can preferably allocate the resources in cloud computing model, the effect of prediction model time is more close to actual time, can efficiently limit the possibility of falling into local convergence, the optimal solution's time of objective function value is shorten which meet the user's needs more.

Keywords: Cloud computing, Network computing, Firefly algorithm, Task scheduling

1. Introduction

Cloud computing is a now wildly used architecture hot, it's product of the development of grid computing, distributed computing, network storage and parallel processing [1]. It shows that the user's applications can operate without personal computer but the server cluster in the Internet. There are three basic forms of cloud computing services including: Infrastructure as a Service (IAAS), Platform as a Service (PAAS) and Software as a Service (SAAS) [2]. In cloud computing, the allocation of resources is a very important issue, the unsatisfactory allocation of resources can easily led the cloud's servers crashed and other servers in idle. So in cloud environment, the problem mostly need to solve is the ways to control any server's resources allocation and use condition by the information communication of local and in the Internet to make better use of the resources. Literature [3] made researches of the resources allocation conditions in different environment. Literature [4] proposed the resources allocation mechanism of self-management, self-adjustment and self-protection. Literature [5, 6] proposed a resources allocation system applies to extensive distributed system, which efficiently increased the system's service quality under cloud computing.

Cloud computing is a combination of parallel computing, distributed computing and virtual technology, a hot technology of nowadays computer industry. The cloud system firstly combined computer, storage device and so on and formed resources pool, then the users could choose the corresponding resources by their needs, this dynamically offers users a computing service environment with reliable and ensure quality of service(QOS). Task scheduling is one of the core technologies of cloud computing which has big effect on the whole performance of cloud computing [7].

Aimed at the task scheduling of cloud computing, a scholar came up with a scheduling algorithm of HADOOP, this algorithm scheduled by task priority and submission time, which was easy to realize but it ignored the difference in tasks and made a long response time of the tasks [8]. Later some scholars proposed the cloud computing task scheduling algorithm with optimal efficiency; the task scheduling algorithm based on cost-driven; the task scheduling algorithm based on trust-driven, those algorithms is flexible and better meet the needs of the users, but they all aimed at a certain object so it's limited in application. A big amount of researches showed, cloud computing used in multiple task scheduling is not only reliable, quality of service(QOS) ensured but also be fair in task allocation, cloud computing multiple task scheduling is a world-recognized NP problem[9]. It's very complex to solve with exhaustive search method, with swarm intelligence algorithm became mature, in recent years, some scholars brought particle swarm optimal algorithm, genetic algorithm and ant colony optimization into the task scheduling of cloud computing, which has a good result . But as the isomerism, dynamics and the differences of users of cloud computing, the cloud task scheduling is a complex problem, single genetic algorithm or ant colony algorithm both has its shortage, and the combinational algorithm based on combination-optimization theory have complementary advantages then find the optimal solution of the problem.

2. Problem Description

2.1. Basic knowledge of cloud model

The cloud model is a transformation model uses linguistic values to express the uncertainty between a certain conception and its quantification expression, it fully combines fuzziness and randomness and forms the mapping between qualitative and quantification, shows in Figure 1.

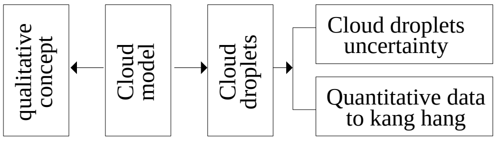


Figure 1. Transformation schematic of cloud mode's qualitative concept and quantitative data

Sets U is discourse domain expressed by accurate numerical value, A is corresponding qualitative concept in U. If quantitative value $x \in u$ and x is a random implementation with likely normal distribution of qualitative concept A in discourse domain U, the certainty degree $A(x) \in [0,1]$ of x to A is also a random number with likely normal distribution, then data array $(x, A(x_i))$ is called as cloud drop, the whole element $x_i\left(i=1,2...n\right)$ in discourse domain U and its certainty degree $A(x_i)$ for A, i,e , n data array $(x, A(x_i))$, forms the cloud model with n cloud drop, calls x distribution in discourse domain U as cloud distribution. The number characteristics of cloud model are expressed as expectation (Ex), entropy (En) and excess entropy (He). Among them, expectation (Ex) refers to the central value of discourse domain U, is the center of qualitative concept, reflects the cloud focus of the whole cloud drop swarm; entropy (En) refers to the range which can be received by fuzzy concept, $En \succ 0$; excess entropy (He) is a uncertain measurement of entropy, i,e , the excess entropy is the entropy's entropy, He >0. The excess entropy reflects the degree of reach an agreement of cloud drop of representation qualitative concept or the concentration degree of cloud drop's representation qualitative concept; the bigger excess entropy is, the qualitative concept has worse common sense or the qualitative concept is worse decentralization.

3. Measurement of Cloud Model's Spray Characteristic

The cloud model's spray characteristic refers to the character of cloud drop distributes around cloud expectation curve's discrete degree. Professor Liuyu, etc. made researches on excess entropy measures cloud drop's discrete degree with fixed entropy. But these works did not show the essence factors of determine cloud model's spray characteristic, i,e , the standard deviation Y's distribution of cloud drop quantitative data X determines the cloud model's spray characteristic. The same as the cloud distribution probability density of cloud model algorithm identified is the theoretical basis of uncertainty reverse cloud model algorithm, this chapter revised the cloud distribution probability density and gave a strict proof according to spray characteristic Y >0.

The positive direction cloud model algorithm steps in one-dimension theory's domain are as following:

Step 1: Generates normal random number y_i whose expectation is E_n , standard deviation is He;

Step 2: Generates normal random number x_i whose expectation is Ex, standard deviation is y_i , x_i is a concrete and quantitative realize of qualitative concept A operates in its corresponding quantitative theory of the domain U, called cloud drop qualitative data;

Step 3: Calculates $r_i = \exp\left(-\frac{(x_i - E_x)^2}{2y_i^2}\right)$, r_i is the certainty degree or subjection degree of x_i belongs to qualitative concept A;

Step 4: Repeats step one to three until generates n cloud.

Prove: because $y - r\left(E_n, H_e^{\ 2}\right)$, En refers to the discourse domain must be greater than zero, as $x \sim N\left(E_x, y^2\right)$, y , as the standard deviation of x , must be greater than zero, so according to normal distribution random variable meets 3σ rule, gets $E_n / H_e \geq 3$. Besides, the probability density of Y is

$$x_i\left(y\right) = \frac{1}{\sqrt{2\pi H_e}} \exp\left[-\frac{\left(t - E_n\right)}{2He^2}\right]$$

When $x_i = y$, the conditional probability density is

$$x_{i,j}\left(x|y\right) = \frac{1}{\sqrt{2\pi y}} \exp\left[-\frac{\left(x - E_x\right)^2}{2y^2}\right]$$

Gets joint probability density through conditional probability density formula:

$$x(i,j) = \frac{1}{2\pi H_e j} - \exp\left[-\frac{\left(j - E_n\right)^2}{2He^2} - \frac{\left(i - Ei\right)^2}{2j^2}\right]$$

Gets probability density which marginal probability density is cloud distribution through joint probability density formula:

$$x_i\left(x\right) = \int_x^y \frac{1}{2\pi Hey} \exp\left[-\frac{\left(y - En\right)}{2He} - \frac{\left(x - E_x\right)}{2y^2}\right]$$

This formula has no analytic form Quod x_i demonstrandum.

From step 2, 3, y is the standard deviation of cloud drop qualitative data X, its distribution character directly determines the cloud drop's distribution character, the bigger distribution scale of Y, the more cloud drop distributes discrete. Because

$$Y \sim N(En, He^2)$$

This text takes $a = En / He$ as the measurement of cloud drop's discrete degree, called spray factor, because qualitative data's standard deviation Y, En and He must be greater than zero at the same time so $a \geq 3$. Spray factor a integrative considers the nature that standard deviation Y of cloud drop's qualitative data X must be greater than zero, the distribution of Y directly affects cloud drop discrete degree and a determines the distribution character of Y, so 0.0 can be the significant digital characteristic of cloud model to presents the discrete condition of cloud drop's distribution. The spray characteristic of cloud model has the following characters:

Character 1: The distribution characteristics of cloud drop's qualitative data standard deviation determines the cloud drop's distribution characteristics, a refers to the cloud drop's discrete degree and $a \geq 3$. The smaller a be, the bigger discrete degree of cloud drop's distribution; when $\alpha = 3$, the discrete degree of cloud drop's distribution reaches the biggest; the bigger α is, the smaller discrete degree of cloud drop's distribution, finally tends to normal distribution. Now the cloud drop all approximate distributes on cloud expectation curve.

Table 1. Sub-tasks and resources table

The sub-tasks	resources	Running time	Running costs	Total resources
n_1	m_1	$t(n_1,m_1)$	$\cos t(n_1,m_1)$	$Sm=\sum_{i=1}^{n} m_i$
n_2	m_2	$t(n_2,m_2)$	$\cos t(n_2,m_2)$	
…	…	…	…	
n_n	m_n	$t(n_i,m_j)$	$\cos t(n_i,m_j)$	

Character 2: cloud distribution's corresponding range of spray factor: $3 \leq a \leq 18$.

The spray factor determines the distribution characters of cloud drop qualitative data, and the kurtosis describe the figure of data distribution at the same time, the kurtosis of normal distribution is 3, if the kurtosis of cloud distribution values around 3, the cloud distribution turns to normal distribution[8]. The kurtosis of cloud distribution defines as following:

Definition 1: the kurtosis of cloud distribution

$$K(X) = 9 - \frac{6}{\left(1+\frac{He^2}{En^2}\right)^2} = 9 - \frac{6}{\left(1+\frac{1}{\alpha^2}\right)^2}$$

7% cloud drop distributes between curve $f_1 = \exp\left[\frac{-(x-Ex)^2}{2(En+3He)^2}\right]$ and $f_2 = \exp\left[\frac{-(x-Ex)^2}{2(En-3He)^2}\right]$. f_1 is cloud mode's outer contour curve, f_2 is inner contour curve.

In cloud computing environment, the mostly used model is Map/Reduce, this model operates well in large-scale parallel task. Especially in cloud computing environment, it needs to processes each cloud user's resource number, time, network channel fee, etc. in time. The currently related task scheduling algorithm focuses on the needs of overall task, considers less about the cloud user's complementing time, which led to unreasonable in time and resources distribution for the users when multiple tasks operates. Supposes cloud client's tasks of cloud computing as table 1:

a) Divides large-scaled task into relatively small tasks, divides in average, the sub-tasks' operating time are similar.

b) The number of resource distribution offers enough for sub-tasks.

c) Reasonable defines sub-task occupies resources time.

N_i refers to the number of sub-tasks, m_i refers to the number of resources, $t(n_i,m_i)$ refers to the time in task i, resources j, $\cos(n_i,m_j)$ refers to the costs in task i, resources j. In these above models, supposes the resources in cloud computing reasonable can be distributes into the computing resources of sub-tasks and ensures the shortest time and the lower costs for complementing the sub-tasks.

4. Proposed Algorithm

4.1. Mathematical model

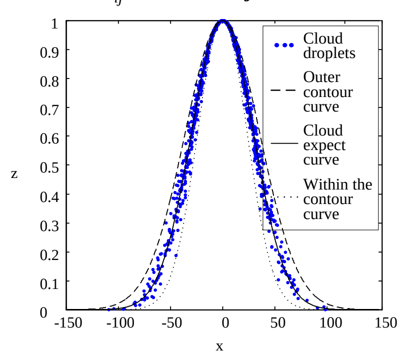
The resources distribution under cloud computing related some specification for cloud users, they have this restrictions as following in general:

Each task can only operate once in one cloud server in the whole machining process.

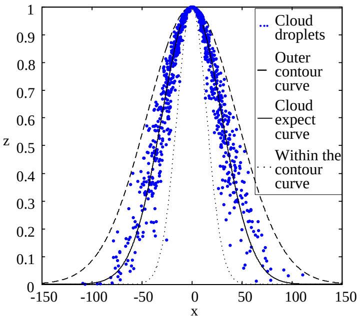
Without considering the superiority of each task.

The mathematic model of this problem can be expressed as following:

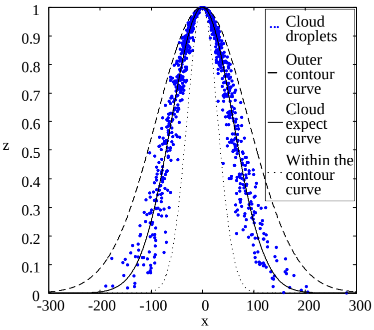
$$FinishTime_{ij} - MachiningTime_{ij} + M\left(1-\alpha_{ijk}\right) \geq FinishTime_{ij} i = 1,2,...n,\ j = 1,2,...n \quad (1)$$



(a) $Ex = 0$ $En = 30$ $He = 2$ $\alpha = 15$



(b) $Ex = 0$ $En = 30$ $He = 5$ $\alpha = 6$



(c) $Ex = 0$ $En = 60$ $He = 10$ $\alpha = 6$

Figure 1. Cloud drop's discrete degree is determined by spray fact

$$FinishTime_{jk} - FinishTime_{ij} + M\left(1-\beta_{ijk}\right) \geq MachiningTime_{oj} i, k = 1, 2, ...n, j = 1, 2, ...n \quad (7)$$

$$FinishTime_{ij} \geq 0, i = 1, 2, ...n, k = 1, 2, ...m \quad (8)$$

$$Min \max\left\{\max FinishTime_{ik}\right\} \quad (9)$$

Formula (6) refers to the operating order of each sub-task determined by each task. Formula (7) refers to the order of each sub-task, formula (8) refers to each sub-task's time variable restriction. Formula (9) refers to objective function. $FinishTime_{ij}$ refers to the complementing time of task i in server k, $MachiningTime_{oj}$ refers to the processing time of task i in server k, M is a coefficient defined values, α_{ijk} and β_{ijk} are expressed as following: α_{ijk} is 1 refers to server j operating task i in server k, α_{ijk} is 0 refers to other conditions. β_{ijk} is 1 refers to task i processing task i in server k, β_{ijk} is 0 refers to other conditions.

4.2. The applications of improved firefly algorithm in cloud computing tasks

In basic firefly algorithm, fluoresce in value judges position the current position of firefly is the best position, when it's in the best position, it can attracts more fireflies to move towards so as to find out the objective value of function, but in the obtaining of fluoresce in, current algorithm is easily to fell into local optimum. Aimed at this, improves formula (5), the fitness value of firefly is between $\left[l_{\min}, l_{\max}\right]$ then gets formula (10)

$$l_j\left(u\right) = l_i\left(u-1\right) + r\frac{l_{\min}}{l_{\max}}\gamma p\left(x_i\left(u\right)\right) \quad (2)$$

The improved fluoresce in value can better avoid felling into local convergence and find best position, this improvement fits the reasonable use of multiple tasks in cloud computing, from formula (6) and formula (7), it ensures the complementing time among tasks in cloud server can reasonable close to processing time and confirms the objective function in formula (9) reaches minimum. Applies improved firefly algorithm into cloud computing, all the processed task uses string encoding, randomly chooses the cloud user's task which needs to be processed, all the tasks are numbered. Each firefly refers is a solution, the firefly's position length refers to all the working procedure. The number of firefly refers to the searching number and space of the solution. The objective function in fireflies turns to the minimum value in cloud computing complementing time.

4.3. Steps of algorithm

a) Initializes each parameter in the algorithm and defines the initial firefly's position, defines the minimum $l_{\min}$ and maximum $l_{\max}$ of fitness degree.
b) Calculates new fluoresce in value by formula (2) and the objective function of firefly is fluoresce in value.
c) Judges the new fluoresce in value by formula (2), controls by formula $r\frac{l_{\min}}{l_{\max}}$.
d) Changes the position of firefly, chooses fireflies which meets the standards by formula (2).
e) Randomly chooses directional firefly i, updating by formula (4).
f) After one iteration, judges whether the iteration meets the finishing condition when enters into next iteration. If not meets, turns to step 2. If satisfied, directly output the optimal solution.

5. Experimental Results

In order to prove the performance of this algorithm, tests in two aspects, one is the performance of algorithm, the other is the task scheduling in cloud computing.

5.1. Performance test of algorithm

The study makes comparison test using three benchmark functions in literature and to test the algorithm's efficiency and performance [8]. Using of MATLAB in Windows.

(a) Sphere function

$$f\left(x\right) = \sum_{i=1}^{m} x_i^2 \quad -100 \leq x_i \leq 100$$

This is a continuous, un-modal convex function, the minimum point of the global function is zero, $i,e,x_i = 0(i=1,2,...n)$, there are on interaction among variables.

(b) Goldstein-Price function

$$f(x)=\left[1+(x_1+x_2+1)^2+(18-13x_1+2x_1^2+5x_1x_2+2x_2^2)\right]$$

$$\left[30+(2x_1-3x_2)^2(19-33x_1+12x_1^2+27x_2^2)\right]$$

$$-3\le 2, i=\{1,2\}$$

$3, i,e,x_i = 3(i=1,2,...n)$, there are on interaction among variables.

(c) Ackely function

$$f(x)=-21\exp\left(-0.2\sqrt{(\frac{1}{n}\sum_{i=1}^{m}\cos 2\pi x_i)+21+r-34\le x_i}\right)$$

This is a multiple hump function with many local minimum points, the minimum point of the whole function is $0, i,e,x_i = 0(i=1,2,...n)$, there are on interaction among variables.

In the setting process of algorithm parameter, the scale of initial firefly is 500, iteration time is 200, parameter of fluoresce in is $p=0.6$, parameter of function is $\gamma=0.4$, initial fluoresce in is $l_0=10$ .The comparison of Sphere function's convergence curves is shown as Figure 2. The comparison of ACKELY function's convergence curves is shown as Figure 3.

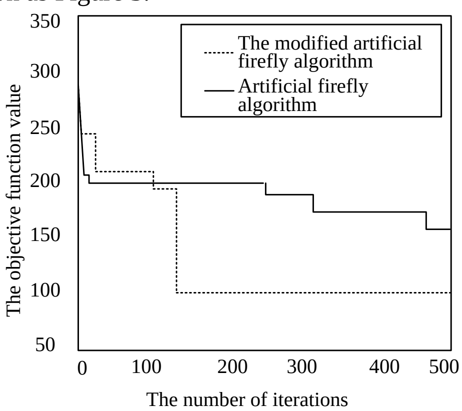


Figure 2. Comparison of Sphere function's convergence curves

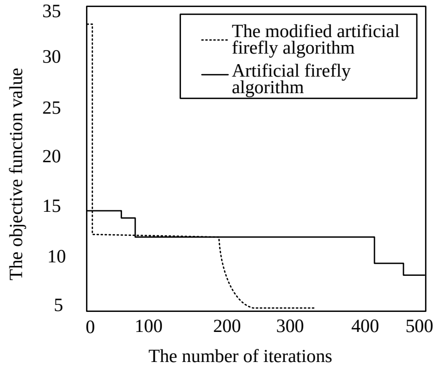


Figure 3. Comparison of ACKELY function's convergence curves

5.2. Task scheduling of improved intelligence firefly algorithm in cloud computing

This text aimed to prove the feasibility and the good stability of the firefly algorithm based on improvement in cloud computing distribution. Takes seven points: A(0,0), B(4,3), C(2,2), D(2,-1), E(3,-4), F(6,-6), G(8,2), node A is client side, B, C, D, E, F node is the server of cloud, G resources side showed as Figure 4. Supposes the resources of client side A and resources side G can bear certain load capacity. AB, AC, AD, AE, AF refers to different tasks the distance refers to the number of required resources. Uses CLOUDSIM [9] platform, T3500CPU and 2GDDR3, Windows XP as operating system, simulates using MATLAB. Sets the number of firefly as 200, node number is 7, prediction transmission time is 100ms, execution time is 300ms, iterations is 400. This text makes comparison of basic firefly algorithm and improved firefly algorithm in cloud computing model, for their number of sub-task and consumed time, showed as Figure 5. The sub-tasks and resources table is shown as Table 2.

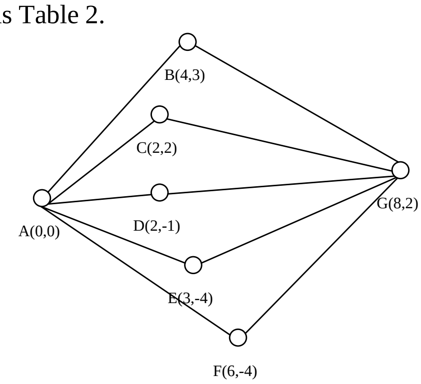


Figure 4. Simulation construction of clouds computing

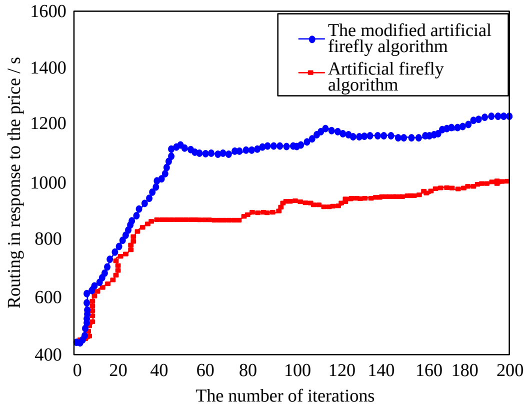


Figure 5. Comparison of the two algorithm's route reply

Table 2. Sub-tasks and resources table

Function	Algorithm	Best solution	Worst solution	Average
Sphere function	intelligence firefly algorithm	0.0212501	0.0298742	0.0255600
	algorithm in this text	0.0019610	0.0105412	0.0062476
Goldstein-Price function	intelligence firefly algorithm	3.0005126	3.0014521	3.0009822
	algorithm in this text	3.0000003	3.0000301	3.0000112
ACKELY function	intelligence firefly algorithm	3.4589131	3.6521469	3.5565298
	algorithm in this text	3.2456321	3.3562153	3.3109179

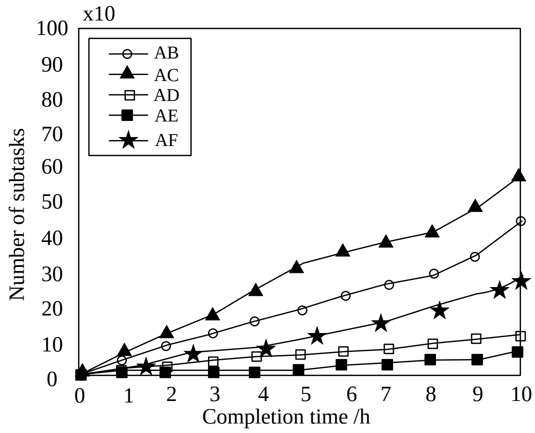


Figure 6. Task and complementing time of modified intelligence firefly algorithm in cloud computing model

From figure 8, with the increasing number of sub-task, the algorithm in this text has certain advantages in task average transmission time and execution time. Mainly because the improved intelligence firefly algorithm can better distributes the resources in the model, prediction model time and effect is gradually close to actual time. From figure 4, the execution time of this prediction method is close to actual complementing time, which meets the task scheduling time in cloud computing environment.

6. Conclusion

How to make full use of the resources in cloud computing environment is a current-focusing problem. The method in this text improved the intelligence firefly algorithm in nature, combined the number of sub-task, resources with algorithm. In intelligence firefly algorithm, improved the method of firefly's fluoresce in position so as to make the firefly find the better object faster. On this improvement, the method reasonable solved the problem of balancing the network load and extending network, enhanced the global convergence of algorithm, it's valuable for increasing network operation. But there are still many practical problems in cloud computing to solve, resources distribution in cloud computing needs to be further researched.

References

[1] Y. Geng, J. He, H. Deng and K. Pahlavan, Modeling the Effect of Human Body on TOA Ranging for Indoor Human Tracking with Wrist Mounted Sensor, 16th International Symposium on Wireless Personal Multimedia Communications (WPMC), Atlantic City, NJ, Jun. 2013.

[2] J. Fan, "The modified Levenberg-Marquardt method for nonlinear equations with cubic convergence", Math. Comput., vol. 81, pp. 447-466, June 2012. http://dx.doi.org/10.1090/S0025-5718-2011-02496-8

[3] S. Li, Y. Geng, J. He, K. Pahlavan,Analysis of Three-dimensional Maximum Likelihood Algorithm for Capsule Endoscopy Localization, 2012 5th International Conference on Biomedical Engineering and Informatics (BMEI), Chongqing, China Oct. 2012 (page 721-725)Provenzi E Gatta C Fierro M et al. A spatially variant white-patch and gray-world method for color image enhancement driven by local contrast[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(10):1757-1770.

[4] Zhou S, Aggarwal G, Chellapa R. Appearance characterization of linear lamebrain object, Generalized photometric stereo and illumination- Invariant face recognition [J]. IEEE Trans on PAMI, 2007, 29(2): 230-245.

[5] A. Ali-Eldin, J. Tordsson, E. Elmroth. A N Adaptive hybrid elasicity controller for cloud infrastructures[C]//IEEE Network Operations and Management Symposium, 2012:204-212.

[6] D. Breitgand, A Epstein. SLA-aware Placement of Multi-Virtual Machine Elastic Services in Compute Clouds[C]//IFIP/IEEE International Symposium on Integrated Network Management. Dublin, Ireland, 2011.

[7] Wen-liang Zhou, Jin Xie, Lian-bo Deng. Sequencing Optimization Method for Drawing Train Operation Diagram of High-Speed Railway, Journal of Multimedia, Vol 8, No 1 (2013), 16-23

[8] Chikara H., Taketoshi K., Norio T., Shigeki K., MotoshiA T., Train stop deployment planning algorithm using a petri-net-based odeling approach quarterly, Report of RTRI, 2009, 50 (1), 8-13. http://dx.doi.org/10.2219/rtriqr.50.8

[9] Young k P, Seokl P, Joong K K. Retinex method based on adaptive smoothing for illumination base on adaptive smoothing for illumination invariant face recognition [J]. Signal Processing, 2008,88: 1929-1945.