**Resource Management in Cloud Computing:**

**Survey**

*Tathagata Mukherjee*

*B.Tech.*

*Department of Computer Science & Engineering*

*Lovely Professional University, Punjab*

*tathagatamukherjee.875@gmail.com*

**Abstract**

There are numerous issues with the current information construction process, which focuses primarily on building networks and management information systems. Cloud computing is the evolution of distributed processing, parallel processing, and grid computing, which makes data saved on the cloud, makes software and services placed in the cloud, and makes it built on top of different standards and protocols, and you can obtain it through all kinds of equipment. This article covers cloud computing and its functions before analyzing the current issues with managing network’s resources and using cloud computing technology and techniques to the creation of a platform for sharing knowledge among servers.

**Keywords:** *Resource management, Replica management, Financial cost, User experience, Edge cloud computing system*

**1.Introduction**

The majority of platforms lack national information technology standards and regular development, which has caused information islands to exist, making it impossible for networks to link to one another between information systems and preventing anyone from sharing resources. Second, there is a clear system management issue. Problems and the management system are the primary influences on information systems, information network center departments, which are in charge of coordinating planning and putting informatization construction into action. However, the majority of its network centers only have two parallel units, which is similar to functional departments. In addition, there aren't any consistent planning permits, there aren't many employees, and the network centers frequently have many laborers on staff. They are typically very busy maintaining systems and equipment, have no more energy, and design the network overall. When this happens, computing clouds will present a better method of managing information resources.

In the computing cloud concept, users can execute and store programs and data storage in huge web server clusters rather than on personal computers, mobile phones, and other devices. Cloud computing models can be used to integrate currently available hardware and software resources, lower the cost of software acquisition, and expedite the construction of application and network information resources while utilizing the limited time and financial resources.

**2.Cloud Computing Technology**

*2.1.Definition and Characteristics of Cloud Computing*

* *Definition of cloud computing:*

Cloud computing is an emerging model, make business computing tasks repository constitute a lot of lay computer resources, so that all application can get computing power, storage space and various software service [1]. The fundamental principle of cloud computing is that user-provided data is stored on the network of data centers, not locally. In order to ensure that these data centers have strong computational power and ample storage space for users, offers cloud computer service management and maintenance. This way, users can access these services from anywhere at any time using any terminal device with Internet access.

* *Characteristics of cloud computing:*

1. Cloud computing provides computing power. In the other side of cloud computing, a large cluster that consists of tens of thousands or more servers provides infinite number of space for the storing and management of data, but also provides infinitely powerful computing capacity for the completion of various application [2].
2. Data storage facilities offered by cloud computing are the most dependable and secure. Information is managed by a qualified team, and the data is stored in the most cutting-edge data center, so users do not need to be concerned about data loss, viruses, or other issues.
3. Cloud computing is easy to utilize and has minimal requirements for client devices. While consumers can access "cloud"-based applications and services using computers and browsers.
4. Utilizing cloud computing makes it simple to share various pieces of equipment and apply data. In the network application model for cloud computing, data is only kept on the other side of the "cloud," the terminal is only connected to the internet, you may access and use the same data, and you can share data with others.

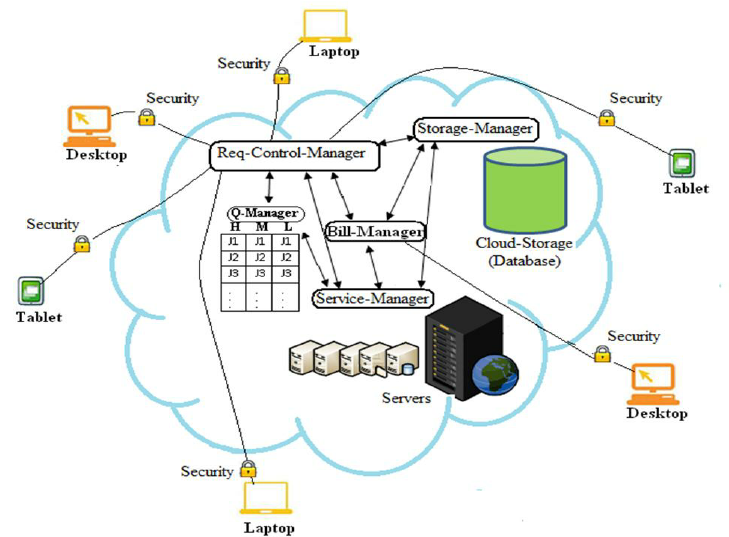


Fig. 1. The organization structure of cloud computing

*2.2.Definition and Characteristics of Edge Computing*

An edge cloud computing is a cloud computing platform based on the core of the cloud computing technology and the edge computing ability, which is built on the edge infrastructure. The edge cloud computing can reduce response latency, pressure and bandwidth cost by managing storage, calculation and data analysis on the network edge. At the same time, it can provide cloud services such as network scheduling and power distribution [8]. Fig. 2 describes the organizational structure of the edge cloud. Because each edge server only has a small amount of storage, an edge cloud also has a small amount of storage. As a result, an edge cloud needs to increase its capacity in order to handle the overload issue when workloads surpass its processing capabilities. In other words, the edge cloud can hire the remote cloud's computation and storage resources to increase its processing power. Due to the dynamic nature of the workloads, there are certain difficulties with resource management.

The cost of rental nodes for resource management is influenced by node configurations. So, an appropriate resource management strategy should be taken into account in order to reduce the financial cost of the rented nodes. The resource management method will be used to rent resources from the remote cloud to fulfill user requests when the workloads exceed the edge cloud's capacity. The edge cloud will release the rented nodes whenever the workloads lessen and it is able to handle user requests. The edge cloud might handle the scenario of a requirement burst and lower the infrastructure cost in this way.

*A diagram of a network

Description automatically generated*

Fig. 2. The organization structure of edge cloud

**Table 1**

The comparison of some resource management strategies

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Ref.** | **Environment** | **Factors influencing resource management** | **Design objective** | **Performance metrics** |
| [9] | Cloud | CPU utilization, memory size, communication delay | The tradeoff between gamer quality of experience and provider net profit | QoE, number of used users, running time |
| [10] | Cloud | Workloads, price of rented resources | The tradeoff between service quality and rented cost | SLA default rate, rented cost |
| [11] | Cloud | Workloads, price of cloud service, effective use of the service | Reducing the rental costs of the resources according to the predicted workloads | Workloads of disk and memory |
| [12] | Cloud | Workloads, price of cloud service | Minimizing the cost of cloud infrastructure service | SLA default rate, number of servers, cost |
| [13] | Cloud | User satisfaction, quality of experience | Minimize the cost cloud service | Average user satisfaction |
| [14] | Multi-cloud | Workloads | Ensuring the availability of the cloud services | Session rate, session concurrency |
| [15] | Edge cloud | Workloads, service cost | Minimizing the cost of edge cloud clusters, while satisfying the cluster workloads | SLA default rate, total cluster cost |

*2.2.1.* *Related work*

The resource and replica management issue has received a great deal of attention in edge cloud computing systems. The research projects are illustrated and discussed in this section in the ways that follow.

* *Resource management strategy*

Resource management can be used to dynamically adjust the cloud resources according to the workload changes [16]. Therefore, many scholars pay attention to the resource management strategy. Table 1 describes the comparison results of some resource management strategies.

Hong et al. [17] proposed the resource management strategy based on the CPU utilization, the memory size, the communication delay and the profit in cloud games. They wanted to achieve the balance between gamer quality of experience and provider net profit. Munoz-Escoi and Bernabeu-Auban [18] presented a survey on looking for the balance between the service quality maximation and the cost minimization of resource provisioning from the remote cloud. Dias et al. [11] designed an elastic provisioning mechanism to reduce the rental costs of the remote cloud resources by forecasting the workloads of the clusters. Han et al. [12] proposed an adaptive scaling algorithm for multi-tier application to reduce the costs incurred by users of cloud infrastructure services. Najjar et al. [13] maximized users’ satisfaction by considering the quality of experience-based resource management approaches, while not extending the extra costs. Paraiso et al. [14] built a multicloud-PaaS architecture to manage elasticity across multiple cloud providers while ensuring the availability of cloud services. Li et al. [15] studied the resource management problem by considering the workloads and the service cost. They aimed to minimize the cost of edge cloud clusters while satisfying the cluster workloads. In this paper, like most of the scholars, we also consider the financial cost of the rented resources of the remote cloud. But, unlike previous research, the penalty function of the rented nodes is designed to improve the financial cost model of the rented nodes and the real-world data set is adopted to evaluate the performance of the proposed resource management strategy. Meanwhile, comparing with our previous works in [15], the differences are described as follows.

1. In [15], the resource management model is built for a single cloud data center. But, in our paper, the multiple data centers are considered providing the rented resources.

1. The data set derived from the Clark Internet Services is used, which is different from the data set in [15].

**Table 2**

The comparison of some replica allocation strategies.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Ref.** | **Environment** | **Factors influencing resource management** | **Design objective** | **Performance metrics** |
| [19] | Cloud | Workloads | Maximizing the data efficiency | Response time, number of distributed transactions |
| [20] | Cloud | Topology of service-based application, communication performance between different computing nodes | Multi-objective optimization of two factors | Makespan, average latency |
| [21] | Cloud | Site, node location, response time | Minimizing response time | Average latency, hit ratio |
| [22] | Hadoop architecture | Power consumption, physical resource waste, file unavailability | Multi-objective optimization of three factors | Power, waste, unavailability |
| [23] | Hadoop architecture | File locality, file availability, replica migration cost | Multi-objective optimization of three factors | Mean network latency, failure rate, mean migration |
| [24] | Edge-cloud | File unavailability probability, node performance, node relative load, budget and deadline constraints, average response time, | Multi-objective optimization of six factors | Mean job time, effective network utilization, storage space utilization |
| [25] | Edge-cloud | File unavailability, node load and network transmission cost | Multi-objective optimization of three factors | Storage space utilization, average response time |

Diagram of a cloud computing diagram

Description automatically generated

Fig. 3. The overview of the resource management strategy and replica allocation strategy in edge cloud computing system.

* *Replica management strategy*

In an edge cloud computing system, with the constant change of node resources, how to allocate the replicas to the appropriate nodes is a key issue. Table 2 describes the comparison results of some replica management strategies. Guo et al. [10] dynamically and adaptively scheduled the data replicas by monitoring the load to achieve the higher scalability, the fault tolerance and the ability to cope with the load. Wu et al. [38] presented a component service replica placement method based on topological matching for a cloud application to improve the performance of service-based applications. With the analysis of the factors that impact the system performance during replica placement, Mansouri and Javidi proposed a hybrid replication strategy to address the problems of the replica placement, the selection, and the replacement to reduce the response time in the large-scale cloud environment [24]. Guerrero et al. [8] determined the virtual machine allocation, virtual machine template selection, and file replica placement by considering the power consumption, the physical resource waste and the file unavailability. The complexity problem of cloud systems and interrelation between their elements were reasonably addressed by achieving the solution of the multi-objective optimization problem. Guerrero et al. [9] studied the replica placement problem based on the file locality, the file availability and the replica migration cost in Hadoop architecture. The multi-objective optimization problem was formulated to achieve an optimal replica placement scheme. Li et al. [18] placed the replicas in the edge-cloud cluster by considering the file unavailability probability, the node performance, the node relative load, the budget and deadline constraints, the average response time and the cluster relative load. Meanwhile, the replica fault-tolerant strategy is designed in [18]. Li et al. [19] built the multi-objective optimization function based on the file unavailability, the node load and the network transmission cost to study the replica placement problem.

*2.3.Key Technologies of Cloud Computing*

Cloud is data intensive supercomputing which regards data as center, show priority in data storage, data management, programming model, and concurrency control and system management etc [3].

* *Massive distributed memory computing:* Cloud computing employs distributed storage to store data, redundant data storage to assure data reliability, high reliability software to replace unreliable hardware, and affordable and dependable systems to ensure high availability, high reliability, and economy.
* *Parallel programming model:* Cloud computing uses the Map Reduce programming model to automatically split the task into multiple sub-tasks, achieve the scheduling and allocation that task is in large-scale computing nodes, and make users easily enjoy the services that the cloud computing brought. This ensures that the background of complex parallel and task scheduling is transparent to users and programmers.
* *Data management technology:* Large data collections are processed, analysed, and effectively served to customers through cloud computing systems. Large data sets must therefore be managed effectively via data management technology. Second, the issue that needs to be resolved is how to extract specific data from a vast data set.
* *Distributed resource management technology:* Distributed resource management systems are used in multi-node concurrent execution environments to guarantee the accuracy of critical system state. When a key node fails, the system requires relocation of services. Distributed resource management technology then coordinates numerous tasks for the usage of resources through the lock mechanism, assuring consistency of data processes.

**3.Network Information Resources**

*3.1.The Problems of Infrastructure*

* *The investment imbalance of hardware and software:*

During the information construction process, paying more attention to the construction of hardware infrastructure, common issues include excessive investment in hardware and insufficient software construction.

* *Duplication of infrastructure:*

All departments have their unique needs for information construction, thus when updating outdated facilities or creating new information infrastructure, a range of servers, switches, and data storage devices must be purchased.

* *The idle of infrastructure:*

After buying many infrastructures, because of various reasons most of time it is in an idle state, which not only improves the operation of equipment costs, but also wastes power resources, capital, and device is waste [4].

*3.2.Integration Problems of Management Information System*

* When implementing information technology, each department uses a different operating system based on their own requirements. This results in poor communication between departments and prevents duties from being assigned to a single point of control.
* Islands of information exist due to a lack of information sharing. Similar to how the early construction processes lacked coordinated planning, its information development processes saw numerous schools, departments, and functional departments operate independently and autonomously, which led to the duplication of information systems and massive waste.
* The information system integration is very difficult. As various departments or branches take different needs on information collection, so different management information systems need to be built. There exists financial management system, educational management system and other management information systems, and these systems are data principles of confidentiality, cannot provide unified interface, make unified system cannot maintain, and requires a lot of professionals and funding, is difficult to connect between the application systems, application system integration is very difficult, to a large part promote the formation of information islands, so that higher level information processing such as data mining, data integration, data acquisition are more difficult, having great impact [5].

*3.3.Management Problems of Information Resources*

* Because of the many software platforms, the department and the operations lack a user interface for unified planning, which frequently necessitates processing. The same user accessing the network with various apps may require numerous landings because to the simultaneous increase in online information resources and applications.
* Information resources lag. Besides internal information resources lack of uniform standards, develop for own and compatibility of system is poor, in their own ways of data information processing, lack of coordination and cooperation [6].

**4.The Application of Cloud Computing**

*4.1.Reducing Cost*

All types of IT resources, including operating systems, servers, routers, memory, etc., will be made available via the cloud for customers, who will only be charged based on demand. The receives assistance from the cloud computing model; instead of investing money in infrastructure construction, they simply purchase virtual cloud infrastructure services from service providers, which significantly reduces the need for such infrastructure construction, operating and maintenance costs, and electricity consumption.

A good solution to this issue will be provided by cloud computing services. In the course of teaching and research, users must install and use a lot of software. Genuine software license fees, subsequent software updates, and the labor-intensive installation of software in the classroom raise new hardware equipment requirements. The initial expenditure can be reduced with cloud computing, high-performance computer equipment is not configured, and the duty of upgrading equipment is delegated to all service providers, the data center constructor, or a supplier of related services. The use of diverse applications is comparable to taking out a loan because they are all sourced from the "sea of clouds" and do not require extensive installation time on every computer. The "cloud" is where updates to software versions and licensing issues are all dumped. Hardware and software resource management and upkeep expenses are drastically reduced, resulting in labor and resource savings.

*4.2. Information Processing*

Now experimental environment is more and more complex, experimental information are more and more, the personal computer is unable to meet the demand of information processing, and cloud computing can utilize a large number of idle computing resources, each computer complete a small amount of computation, large amount of computing are integrated to speed up the completion of the experiment time and improve efficiency. The departments will make all data stored in the database in the cloud, using cloud computing, unified platform for data storage, data management and maintenance, thus reducing the cost of data maintenance [7], to a certain extent solving the problems that the increasing amount of data make the expansion for database servers and storage equipment.

Users can easily access their important data using the cloud storage services of cloud computing, which can provide security for critical data protection. They do not have to worry about data loss due to storage equipment damage or virus attacks because these are handled by professional service providers that can offer cloud storage and have cutting-edge technology. In the meanwhile, centralized data storage makes it simpler to execute safety monitoring because the data is kept in one or more data centers. The managers are in charge of resource allocation, load balancing, software deployment, security, control, and efficiently lowering user costs. They are also capable of managing all data.

*4.3.On-line Office System*

The user can only access the server through a browser, aware that all programs and data are on the server side. This includes the ability to modify forms, use a courseware editor, and edit files directly. First of all, network security performance has increased, files and applications are stored on the server, making it simple to supervise network management employees, respond immediately to illegal intrusions, and stop unintended unauthorized copying of personnel files; Second, cloud computing decreases document duplication, frees up storage, and allows for the storage of all types of resources; third, the cost is minimal. The client PC can be fairly basic because the system is server-side.

*4.4.Cloud Computing Security Platform.*

Now that viruses and Trojans are spread via networks, cloud computing environments with grid computing and deployed honeypots everywhere can quickly respond to newly emerging attack types, update their rule bases to provide a strong defense, significantly improve their virus sample collection, and shorten the response time to threats.

*4.5.Resource Sharing*

The scalability of cloud computing is quite great; different institutions can place already-existing hardware resources in the "cloud", reducing the time and money invested by individual institutions and realizing resource sharing. The management of institutions must get ready as soon as possible so they may benefit from the flexibility and efficiency that cloud computing security services can offer as the technology and applications for cloud computing become more developed.

**5.Conclusion**

In conclusion, this paper has highlighted the critical issues within the current information construction process, emphasizing the challenges related to infrastructure, integration of management information systems, and resource management. It has also introduced cloud computing as a transformative technology that offers solutions to these problems by providing cost-effective, scalable, and secure computing resources. Cloud computing's characteristics, such as on-demand access, data reliability, and ease of use, have been discussed, demonstrating its potential to revolutionize the way information resources are managed and shared. The paper has also delved into the emerging field of edge cloud computing, which further enhances the capabilities of cloud technology by reducing response latency and improving resource management.

**6.Future Works**

As we look towards the future, there are several avenues for further research and development. First and foremost, there is a need for more comprehensive studies on the practical implementation and adoption of cloud computing in various sectors, including education, research, and business. Additionally, ongoing research in edge cloud computing should explore advanced resource management strategies to address the challenges posed by the dynamic nature of workloads. Furthermore, ensuring the security of cloud computing platforms, especially in the face of evolving cyber threats, remains a pressing concern. The development of robust security solutions for cloud environments is an essential area for future research. Finally, the paper hints at the potential for multi-cloud approaches, which could offer even greater flexibility and redundancy, making it vital to investigate how to efficiently manage and coordinate resources across multiple cloud providers. In this ever-evolving landscape, future work in cloud computing will play a crucial role in shaping the information infrastructure of tomorrow.

**References**

[1] Liu Peng,The characteristic and definition of cloud computing,in press.

[2] Li Kai Fu,The walking of cloud-The coming of cloud computing,in press.

[3] Yang kerong,Han xingshun,Cloud computing and the solution of information construction problem in colleges and universities,Cornputer knowIedge and Technology, vol5, pp:5571-5573, July 2009.

[4] Wang yang,Cloud computing in the college of education application,2009,4.

[5] Jim Baty,Take your business to a higher level,Sun Cloud Computing[J],pp:45-48,2009.

[6] Mike Hogan,How databases can meet the demands of cloud computing, Sun Cloud Computing [J],2008ˊ

[7] Sun shao leng,Luo zhiguo,Xu meng,The research and implementation of clouding computing,Telecom Engineering Technics and Standardization,pp:23-27,April 2009.

[8] J. Pan, J. Mcelhannon, Future edge cloud and edge computing for internet of things applications, IEEE Internet Things J. 5 (1) (2018) 439–449.

[9] I.A.T. Hashem, I. Yaqoob, N.B. Anuar, S. Mokhtar, A. Gani, S.U. Khan, The rise of “big data” on cloud computing: review and open research issues, Inf. Syst. 47 (2015) 98–115.

[10] F.D. Munoz-Escoi, J.M. Bernabeu-Auban, A survey on elasticity management in PaaS systems, Computing 99 (7) (2017) 617–656.

[11] A.D.S. Dias, L.H.V. Nakamura, J.C. Estrella, R.H.C. Santana, M.J. Santana, Providing IaaS resources automatically through prediction and monitoring approaches, in: IEEE Symposium on Computers and Communications, Funchal, Portugal, 2014, pp. 1–7. 23–26 June.

[12] R. Han, M.M. Ghanem, L. Guo, Y. Guo, M. Osmond, Enabling cost-aware and adaptive elasticity of multi-tier cloud applications, Fut. Gener. Comput. Syst. 32 (2014) 82–98.

[13] A. Najjar, C. Gravier, X. Serpaggi, O. Boissier, Modeling user expectations & satisfaction for saas applications using multi-agent negotiation, in: IEEE/WIC/ACM International Conference on Web Intelligence, Omaha, NE, USA, 2016, pp. 399–406. 13–16 Oct.

[14] F. Paraiso, P. Merle, L. Seinturier, Managing elasticity across multiple cloud providers, in: International Workshop on Multi-cloud Applications & Federated Clouds, Prague, Czech Republic, 2013, pp. 53–60. 22-22 Apr.

[15] C. Li, H. Sun, Y. Chen, Y. Luo, Edge cloud resource expansion and shrinkage based on workload for minimizing the cost, Fut. Generat. Comput. Syst. 101 (2019) 327–340.

[16] Y. Aldhuraibi, F. Paraiso, N. Djarallah, P. Merle, Elasticity in cloud computing: state of the art and research challenges, IEEE Trans. Services Comput. 11 (2) (2018) 430–447.

[17] H.J. Hong, D.Y. Chen, C.Y. Huang, K.T. Chen, C.H. Hsu, Placing virtual machines to optimize cloud gaming experience, IEEE Trans. Cloud Comput. 3 (1) (2015) 42–53.

[18] F.D. Munoz-Escoi, J.M. Bernabeu-Auban, A survey on elasticity management in PaaS systems, Computing 99 (7) (2017) 617–656

[19] W. Guo, X. Wang, Y. Don, A workload-based dynamic adaptive data replica placement method, in: 11th Web Information System and Application Conference, Tianjin, China, 2014, pp. 184–187. 12–14 Sept..

[20] J. Wu, B. Zhang, L. Yang, P. Wang, C. Zhang, A replicas placement approach of component services for service-based cloud application, Cluster Comput. 19 (2) (2016) 709–721.

[21] N. Mansouri, M.M. Javidi, A hybrid data replication strategy with fuzzy-based deletion for heterogeneous cloud data centers, J. Supercomput. 74 (10) (2018) 5349–5372.

[22] C. Guerrero, I. Lera, B. Bermejo, C. Juiz, Multi-objective optimization for virtual machine allocation and replica placement in virtualized Hadoop, IEEE Trans. Parallel Distrib. Syst. 29 (11) (2018) 2568–2581.

[23] C. Guerrero, I. Lera, C. Juiz, Migration-aware genetic optimization for MapReduce scheduling and replica placement in Hadoop, J. Grid Comput. 16 (2018) 265–284.

[24] C. Li, Y.P. Wang, Y. Chen, Y. Luo, Energy-efficient fault-tolerant replica management policy with deadline and budget constraints in edge-cloud environment, J. Netw. Comput. Applic. 143 (2019) 152–166.

[25] C. Li, Y.P. Wang, H. Tang, Y. Luo, Dynamic multi-objective optimized replica placement and migration strategies for SaaS applications in edge cloud, Fut. Generat. Comput. Syst. 100 (2019) 921–937.

[26] X. Wang, S. Yang, S. Wang, X. Niu, J. Xu, An application-based adaptive replica consistency for cloud storage, in: 2010 Ninth International Conference on Grid and Cloud Computing, Nanjing, China, 2010, pp. 13–17. 1–5 Nov.

[27] J. Wu, B. Zhang, L. Yang, P. Wang, C. Zhang, A replicas placement approach of component services for service-based cloud application, Cluster Comput. 19 (2) (2016) 709–721.

[28] J. Yang, C. Liu, Y. Shang, B. Cheng, Z. Mao, C. Liu, L. Niu, J. Chen, A cost-aware auto-scaling approach using the workload prediction in service clouds, Inf. Syst. Front. 16 (1) (2014) 7–18.

[29] Z.H. Zhan, X.F. Liu, Y.J. Gong, J. Zhang, H.S.H. Chung, Y. Li, Cloud computing resource scheduling and a survey of its evolutionary approaches, ACM Comput. Surv. (CSUR) 47 (4) (2015) 63:1-63:33.

[30] Q. Zhong, J. Hao, L. Yang, Y. Yao, A replica consistency detection method for text images based on multi-feature assessment, in: 2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings, Taipei, China, 2016, pp. 1–4. 23–26 May.