ISSN 2250-3153

Survey of Various Approaches on Video Recognition Systems

Anjitha V*, Gnana Sheela K **

* Department of Electronics and Communication Engineering,TIST,CUSAT,Kerala,India
** Department of Electronics and Communication Engineering,TIST,CUSAT,Kerala,India

Abstract- Video-based object or face recognition services on mobile devices have recently gained significant attention.Cloud systems start to be utilized for services to analyze user data in the region of computer vision. In these services, key points are extracted from images or videos and the data is identified by machine learning with large database of cloud.Detecting Person-Of-Interest (POI), e.g., fugitives, criminals and terrorists in public spaces is a critical requirement of many law enforcers and police officers. Face recognition applications for airport security and surveillance can benefit from the collaborative coupling of mobile and cloud computing as they become widely available today.

Index Terms- Cloud Server, Key Point of Interest, Scale-Invariant Feature Transform

I. Introduction

Potential of a cloud system combining a smart device and cloud servers is increasing dramatically. One of the examples is "Siri" which offers a friendly web knowledge navigator based on natural language user interface. Since latest portable devices equipped not only a microphone but also a high resolution camera, this kind of cloud based framework is also promising to create various kinds of video based recognition systems.

II. LITERATURE SURVEY

Moravec et al (1981) developed a corner detector for image matching, but it is less repeatable and is only 20% accurate.The Moravec detector was improved by Harris and Stephens (1988) to make it more repeatable for small image variations and near edges. The Harris corner detector is very sensitive to changes in image scale, so it does not provide a good basis for matching images of different sizes. Harris corner detector has been widely used for image matching tasks. While these feature detectors are usually called corner detectors, they are not selecting just corners, but rather any image location that has large gradients in all directions at a predetermined scale. The drawback of this system is it is very sensitive to changes in image scale and does not provide a good basis for matching images of different sizes. It is only 50% accurate.

Lowe D G et al (2004) developed a method for extracting distinctive invariant features from images that can be used to perform reliable matching between different views of an object or scene.The features are invariant to image scale and rotation, and are shown to provide robust matching across a substantial range of affine distortion, change in 3D viewpoint, addition of noise, and change in illumination. The features are highly distinctive, in the sense that a single feature can be correctly matched with high probability against a large database of features from many images. This paper describes an approach to using these features for object recognition. The recognition proceeds by matching individual features to a database of features from known objects using a fast nearest-neighbor algorithm, followed by a Hough transform to identify clusters belonging to a single object, and finally performing verification through least-squares solution for consistent pose parameters.

This approach to recognition can robustly identify objects among clutter and occlusion while achieving near real-time performance.The described SIFT key points are particularly useful due to their distinctiveness, which enables the correct match for a key point to be selected from a large database of other key points [1]. The key points are shown to be invariant to image rotation and scale and robust across a substantial range of affine distortion, addition of noise, and change in illumination.Paul A K et al (2006) presents multiclass object classification and recognition using smart phone and cloud computing (client server) technology. The image recognition is scale and rotation invariant and the speed of operation is very fast. The response time from the server is very fast. This system can be extended for navigation applications and landmark detection, driver assistance. It isapplied for recognizing two dimensional texture based object.Using SIFT the average classification rate is 84%. The computing power of the smart phone is not enough to give the prompt response.

Heinzelman et al(2011) proposed the design and implementation of face recognition applications using mobile-cloudlet-cloud architecture named MOCHA. This architecture is designed to minimize the overall response time of the face detection and face recognition algorithms given heterogeneous communication latencies and compute powers of cloud servers at diverse geographical placements. The obtained results are more intelligent task partitioning algorithms employed by the cloudlet permits response-time improvement by offloading work from the mobile device, 2) the response time decreases as the number of cloud servers increase and this improvement is more emphasized when cloudlets are in place, and 3) communication latencies affect the response time considerably, which can be partially coalesced when cloudlets are used as buffers. It is complex, and is only 69% accurate.

Yuan J et al (2012) proposed a method for local occupancy information based on the local 3D point cloud around each joint using fouriertemporal pyramid features. The disadvantages are,it has high-ordering features and complicated learning procedures. It has 85.7% recognition rate.Jong et al (2012) describes an expert system of bio information, which is combined with the smart devices using wireless sensor network (WSN). The HRV detection parameter value is adopted the criteria and basis for the features of diabetes by using artificial neural network (ANN) algorithm. The accuracy of ANN diagnose has 87.33%A general purpose, low-cost, wide area, remote monitor, mobile and real-time analyzer is designed and implemented for investigation of the heart rate variability. Zigbee is seen that short distance communication within a care zone.

Prakashet al (2013) investigates a wearable computerizedeyewear based face recognition system. This system is a portable device which can accompany a police officer during patrolling or other tasks. The eyewear is connected to a cloud based face recognition system via wireless networks. Facial images captured by the mounted camera are sent to the cloud for identity retrieval. When the system finds POI, it would alert officers via overlaying a virtual identity tag on the real POI's face on the transparent screen of the eyewear. We provide approaches to greatly minimize recognition time, including leveraging the large storage and high computational capacities provided by the cloud. The cloud enables nationwide POI database and supports parallel computing for face recognition.

Gook et al (2013) proposed a system using cloud computing, image processing, information mapping and data synchronization. The advantage of the system is better and faster method for image data synchronization. It is most efficient way with using cloud server included human tracking at the end results .Even though it has 334.033 millisecond of fastest execution time and faces several security problems.Yiannis et al (2014) propose a contextual bandit framework for learning contention and congestion conditions in object or face recognition via wireless mobile streaming and cloud-based processing .main advantages are it maximize the average recognition rate and efficiently allocating transmission settings and has highly-varying contention levels in the wireless transmission, as well as the variation in the task-scheduling congestion in the cloud .But it is complex. The recognition accuracy rate per transaction is 90%.

Althloothi et al (2014) designed method with multi-fused features (i.e., motion of 3D joint positions and shape information) along with multiple kernel functions achieves 93.1% recognition rate.This method uses human activity recognition (HAR),depth-based life logging HAR system , depth imaging sensor and hidden Markov Models(HMM). The advantage of this system is that,once the life logging HAR system is trained, the system can be activated to produce life logs by recognizing the learned human activities.Performance of the proposed life-logging HAR technique, achieving mean recognition rates of 92.33%, 93.58% and 90.33% over the conventional methods having PC features as 78.42%, 77.83% and 72.25% and IC features as 82.58%, 82.25% and 78.33% in smart indoor environments.

Takahiro et al(2014) discussed low complexity SIFT (Scale-invariant feature transform) based key point extraction algorithm and its hardware engine capable of operating at full-HD 60fps video .As a technique to reduce network bandwidth, a keypoint of interest (KOI) detection algorithm based on spatio-temporal feature considering mutual dependency and camera motion is also discussed.The proposed algorithm achieves about 95% reduction of key points and 53% reduction of computational complexity. The reduction of descriptor data communicated with cloud systemsand reduction of computational complexity of descriptor calculations are achieved .It reduces noise by comparing with states of surrounding key points.

III. COMPARATIVE ANALYSIS

Table 1. Comparative analysis of various approaches

No	Author	Year	Methodology	Advantage	Disadvantage	Result
1	Ikenaga T et al	2014	1.A low complexity SIFT (Scale-invariant feature transform) 2.Kkey point of interest (KOI) detection algorithm KLT tracker	1.Reduction of descriptor data communicated with cloudSystems 2. Reduction of computational complexity of descriptor Calculations 3. Reduction of data amount of key points 4.It reduces noise by comparing with states of surrounding key points.	Proposal detects key points from only human which moves largely and outstanding texture whose gradient is large	The proposed algorithm achievesabout 95% reduction of key points and 53% reduction of computationalcomplexity.
2	Jalal A et al	2014	1.Human activity recognition (HAR) 2.Depth-based life logging HAR system 3. Depth imaging	1.Once the life logging HAR system is trained, the system can be activated to produce life logs by recognizing the	1.complex	performance of the proposed life-logging HAR technique, achieving mean recognition rates of 92.33%, 93.58% and 90.33% over

			sensor 4.Hidden Markov Models	learned human activities		the conventional methods having PC features as 78.42%, 77.83% and 72.25% and IC features as 82.58%, 82.25% and 78.33% in smart indoor environments
3	Althloothi et al	2014	Multi-fused features (*i.e.*, motion of 3D joint positions and shape information) along with multiple kernel functions		1.Large-dimensionality features which may be unreliable for postures with self-occlusion 2. Needs high computational times that make it impractical for long-term HAR and real-time applications.	93.1% recognition rate
4	Atan O et al	2014	1.multi-user multi-armed bandits 2.computing cluster	1.Maximize the average recognition rate 2.Efficiently allocating transmission settings 3. Highly-varying contention levels in the wireless transmission,	1.Set of device transmission settings grows combinatorial 2.complex	Same recognition accuracy rate per transaction (90%)

5	Gook L B et al	2013	1.Cloud Computing 2.Image Processing 3.information mapping 4.data synchronization 5. FAST Features Detection 6 .SURF Features Detection	1.A better and faster method for image data synchronization 2.Most efficient way with using cloud server included human tracking at the end results 3. Better computation speed	1.Security problems	334.033 millisecond of fastest execution time
6	Wang X et al	2013	computerized-eyewear POI detection	1.Rapid detection of POIs from a large POI database	1.No information on the evaders (*e.g.*, gender, race or height)	Recognition time is reduced from 33748*ms* to only 202*ms*, i.e., 99*:*4% of time is saved.
7	Yuan, J et al	2012	local occupancy information based on the local 3D point cloud around each joint, Fourier temporal pyramid features	efficient	1.high-ordering features 2.complicated learning procedures,	85.7% recognition rate
8	Huang C H et al	2012	1.wireless sensor network (WSN) 2.ANN algorithm 3.heart rate variability (HRV)	1.A general purpose, low-cost, wide area, remote monitor, mobile and real-time analyzer 2.A reference to the medical staff at the time of diagnosis	1.Training is required for data sets	Accuracy of ANN diagnose has 87.33%

9	Muraleedharan R et al	2011	1.Mobile-cloudlet-cloud architecture named MOCHA	1.Minimize the response time given diverse communication latencies and server compute powers. 2.Designed to minimize the overall response time of the face detection and face recognition algorithms 3. Technically feasible 4. cost-effective	1.Complex architecture	69% face recognition
10	Paul A K et al	2006	1.Cloud computing 2.Scale Invariant Feature Transforms (SIFT)	1.Image recognition is scale and rotation invariant 2.Speed of operation is very fast 3.Response time from the server is very fast	1.Computing power of the smart phone is not enough to give the prompt response	Average classification rate is 84%.
11	David .G Lowe et al	2004	1.SIFT(Scale Invariant Feature Transform) 2.Hough transform	1.Distinctiveness, 2.It is invariant to image rotation and scale and robust across a substantial range of affine distortion,addition of noise, and change in illumination 3. computation is efficient 4. Resistant to even large amounts of pixel noise,	1.,Major cause of error is the initial location and scale detection.	Recognition accuracy for 3D objects rotated in depth by 20 degrees increased from 35% for correlation of gradients to 94%

12	Harris et al	1988	Corner detection	1.More repeatable undersmall image variations and near edges 2.They are not selecting just corners, but rather any image location that has large gradients in all directions at a predetermined scale.	1.Very sensitive to changes in image scale 2. Does not provide a good basis for matching images of different sizes	50% accurate
13	Moravec et al	1981	corner detector	Useful for image matching	Less repeatable	20% accurate

Table 1 shows comparative analysis of various recognition systems. By analyzingexisting methods it is clear that proposed method is more effective.

IV.CONCLUSION

This survey presented that the smart feature detection device for cloud based video recognition system is more effective than other techniques of recognition systems. Because it describes a low complexity SIFT based key point extraction algorithm and its hardware engine. It also introduces key point of interest detection algorithm based on spatio-temporal feature. Both are key to implementing a smart feature detection device and will make a great contribution to increase the potential for a cloud based video recognition system.

It's advantages are reduction of descriptor data communicated with cloud Systems,reduction of computational complexity of descriptor Calculations and reduction of data amount of key points.

REFERENCES

[1] D. G. Lowe,“ Distinctive image features from scale-invariant keypoints,”, Int.Journal of Computer Vision, 60, pp. 91-110,2004.Takahiro Suzuki, Takeshi Ikenaga, "Keypoint of Interest Based on Spatio-temporal Feature Considering Mutual Dependency and Camera Motion", The Sixth International Conferences on Advances in Multimedia (MMEDIA 2014),Feb. 2014.

[2] Xi Wang, Xi Zhao, Varun Prakash, Zhimin Gao, Tao Feng, Omprakash Gnawali, Weidong Shi “Person-Of-Interest Detection SystemUsing Cloud-Supported Computerized-Eyewear”

[3] AhmadJalal 1, Shaharyar Kamal 2 and Daijin Kim,“A Depth Video Sensor-Based Life-Logging Human Activity Recognition System for Elderly Care in Smart Indoor Environments”ISSN 1424-8220 ,www.mdpi.com/journal/sensors,2014

[4] Chen-Shen Huang, Kong-Sheng Huang, Gwo-Jia Jong,”Artificial Neural Network Integrated Heart Rate Variability with Detection System,2012

[5] Tolga Soyata, Rajani Muraleedharan-Sreekumaridevi, Colin Funai, Minseok Kwon, Wendi Heinzelman,"Cloud-Vision: Real-time Face Recognition Using aMobile-Cloudlet-Cloud Acceleration Architecture",2011

[6] Onur Atan, Yiannis Andreopoulos"Bandit Framework For Systematic Learning InWireless Video-Based Face Recognition"

[7] G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.

[8] W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

[9] H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1985, ch. 4.

[10] B. Smith, "An approach to graphs of linear forms (Unpublished work style)," unpublished.

[11] E. H. Miller, "A note on reflector arrays (Periodical style—Accepted for publication)," *IEEE Trans. Antennas Propagat.*, to be published.

[12] J. Wang, "Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication)," *IEEE J. Quantum Electron.*, submitted for publication.