Use of Everyday Mobile Video Cameras in IoT Applications

Nikita A. Bazhenov, Dmitry G. Korzun
Petrozavodsk State University (PetrSU)
Petrozavodsk, Russia
{bazhenov, dkorzun}@cs.karelia.ru

***Abstract*—Nowadays there is a need to receive and process large video streams in shortest time. This can be realized by interaction of video cameras with additional components of the environment (IoT-devices). Thus, third-party devices will be involved, allowing to simultaneously calculate large amounts of data. Applied methods and calculations will be intelligent due to the increase effectiveness of video information management and the use of other peripheral devices for processing. Intellectualization technology can be used in many applied areas. So, in medicine, you can monitor the patient at home using reflective elements (Parkinson's disease). Vision systems allow you to recognize different information based on video data. In the industry, monitoring tasks arise production equipment directly by the employee (for example, identification of deviations in the work).**

I. PROBLEM

The recent advances in digital video technology have provided a large collection of video data. The retrieval and browsing such a large collection of video data is a time consuming and tedious task. Moreover, processing this huge collection of video data demands a lot of resources like man power, storage, hardware and software. It is necessary to have a practical and efficient approach to understand the semantics of video clearly and quickly. An efficient summary of video facilitates users to grasp the important information quickly without watching the whole video. There are two types of summarized videos, using still key frames and video skims. The former type is simply a collection of prominent images from a video. Video skims, also known as dynamic video segments, is the collection of consecutive frames extracted from a video [1].

Every day mobile video cameras can be used as form of sources to receive video streams. It can be expensive, have a complicated and managable interface. And it can be a laboratory, street, at-home, everyday video cameras. The main parameter is to obtain a video stream, which is then converted into one large video.

Services are built and are delivered anywhere and at any time using the surrounding devices of our everyday life. In article [2] video tracking capabilities at home for the patient are being discussed where there are medical problems and the need for an active study of patients with Parkinson's disease at home. First, the problem of monitoring the movement of video in patients with PD is determined [3]. Next, methods that support the Internet for traffic video tracking, which are limited to professional settings of the medical environment. Finally, it is possible to create a personal home laboratory based on such cheap home cameras, like in any smartphone. Early experiment showed that such The cameras provide reliable capture quality for practical use in PD monitor the movement of the patient.

As example there can also be a systems of joint activity (teamwork) in which humans take place in a mutually beneficial situation. Communication plays a leading role in building up services based on coordination and collaboration. In a simple case, people, for example, can try to solve the problem by placing all their faces on one video image in such a way that the camera can recognize all of their faces. In a difficult case, several people using the camera and trying to recognize complex objects, learning their trajectories, behavior and image of motion (using graphs as example).

Emergence of smart spaces, where several devices can use a common view of resources and services, characterize the modern usage of devices. The model based on smart spaces using mobile devices and everyday video cameras can be built. The paper presents a reference model that provides a link between smart space devices (using Semantic Information Broker, SIB) and usage of smartphones.

Modern device usage is moving towards so called smart spaces where a number of devices can use a shared view of resources and services. Smart spaces can provide better user experience by allowing a user to bring in new devices flexibly and to access all the information in the multi device system from any of the devices. As is assumed by such environment, one of the essential features is the information sub-system that should provide a permanent robust infrastructure to store and retrieve the information of different types from the multitude of different environment participants.

II. SYSTEM DESIGN

Smart spaces support development of advanced service-oriented applications that introduce intelligence into Internet of Things environments. Current development meets the performance challenge, since the intelligence is achieved by the cost of performance. Study was a representative application for smart spaces - the SmartRoom system. It assists humans in such collaborative work activity as conferences, meetings, and seminars. Using the existing software Smart-M3-based implementation, we can evaluate the performance of the three important components of service construction and delivery within a localized IoT–environment of the room. This conclusions provide performance bounds convenient for a wide class of Smart-M3 applications [4].

ISSN 2305-7254

Smart-M3 is an open-source platform that implements the smart space concept and maintains its content with ontological representation. SmartSlog is a tool for developing Smart-M3 knowledge processors (KPs) in terms of OWL classes, properties, and individuals. It describes new schemes of KP generating process and ways for improving KP development based on ontologies integration and general-properties. New generating schemes use Protege and ontologies meta-data.

Protege is an ontology editor that is used to build the metadata. The meta-data is used with SmartSlog generator as an extra information for KP generation. The ontology integration allows KP to work with knowledge represented by several ontologies in the smart spaces.

In our case, there can be an application that will be a service (a set of services), working on the basis of smart spaces (with use of SIB). As example: integration and writing a separate module for SmartRoom (system supporting collaboration activities localized in a room: a set of digital services is available for organizers and participants).

It is necessary to identify entities with which our service will interact:

1) Cameras
2) Video server
3) Clients (users)

What information should be stored? What should SIB know?

Cameras –>Video Server: cameras must transmit the video stream to the video server. It is not necessary for cameras to know which Video Server to send video data from it, it is the work of the video server.

Video Server –>Cameras: video server must have information about the cameras that transmit video streams to it. For each camera in databases should be given information about it and its place to store video streams.

Cameras –>Clients: cameras transmit video stream to clients. In the case of weak and cheap cameras, it is necessary to observe a limit of simultaneously connected users to support stable work.

Clients –>Cameras: for convenience and ease of use, it is necessary that clients can connect to the camera in the fewest possible steps. Information about early connections and interactions should be preserved and displayed to the client when the service is started.

Video Server –>Clients: video server must know which clients have access to provide streams. This can be implemented with the help of authorization. Thus: some number of users are tied to a certain number of cameras. One client can have access to many cameras. Multiple clients can connect to one camera.

Clients –>Video Server: it is logical that the video server, upon request of the client, provide recorded streams from video cameras. It is possible that client will want to see the records even from the time when he did not use the service. Therefore, video server after the authorization must provide a variety of information.

It is necessary to store so many connections. In a simple case, all this can be stored on the client's smartphone. But, because of cheap smartphones that are used, it is problematic to store. Therefore, links can be stored ”shortened” that are stored, for example, 1 day and provide not all information but only the most necessary.

Bounds can also be stored on a router or microcomputer (for example, Raspberry Pi). It is useless to select a separate server for storing bounds. And it makes sense when, in addition to the links, various algorithms for recognizing images from a video stream, processing images from several cameras, etc. will be used (possible algorithms are described further).

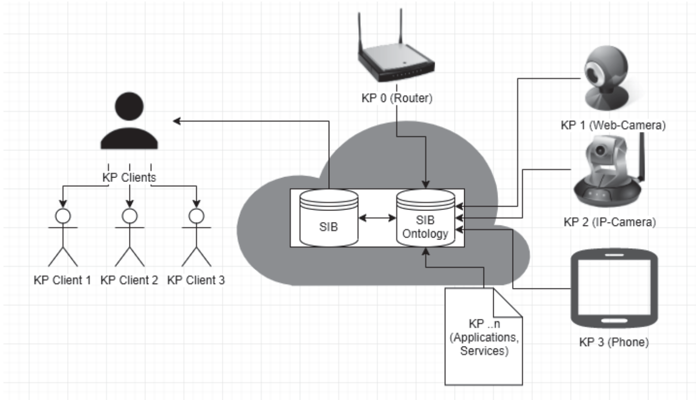

Fig. 1. Videocameras in smart spaces

We propose to allocate two entities for storing bounds in smart spaces (that is shown in Fig. 1). SIB - the provision of ready information about relationships between entities, appearance. SIB Ontology is to generate and store connections between entities.

There are can be some requirements: In some environments, it is more difcult for distributed systems to cooperate. In fact, some distributed systems are highly heterogeneous and might not readily cooperate. In order to alleviate these problems, we have developed an environment that preserves the autonomy of the local systems, while enabling distributed processing. This is achieved by: modeling the different application systems into a central knowledge base (called a Metadatabase); providing each application system with a local knowledge processor; and distributing the knowledge within these local shells. The knowledge decomposition process used for its distribution was described. The decomposition process is used to minimize the needed cooperation among the local knowledge processors, and is accomplished by serializing the rule execution process. A rule is decomposed into an ordered set of subrules, each of which is executed in sequence and located in a specic local knowledge processor. The goals of the decomposition algorithm are to minimize the number of subrules produced, hence reducing the time spent in communication, and to assure that the sequential execution of the subrules is equivalent to the execution of the original rule.

We consider existing solutions for using smartphones in PD people monitoring, assessment, and management. In particular, we utilize a novel remote approach to enrollment, in which participants self-guide through visually engaging yet complete informed consent process prior to deciding to join the study. A

TABLE I. WHAT SHOULD SIB KNOW?

-	Cameras	Video Server	Clients
Cameras	-	Transmitting video stream	Limiting the number of clients
Video Server	Informating about cameras	-	Connecting clients using limitations
Clients	Accessing to a several cameras	Requesting for records	-

critical aspect of this transparent consent process is providing the participants with an explicit decision point specifying if the data they donate to the study can also be used for secondary research purposes. In the experiments, special tools are used to support developing the motor skills and mental abilities of the PD patients.

Generating video data into a fast-response transmission and high-resolution video using cost-effective video processing is desirable in many applications, including Internet of Things (IoT) applications [5]. Due to the rapid development of IoT intelligent sensor applications in real time, it required processing a huge amount of data for the video signal, which led to video compression technology.

For the simplest case,we used only a smartphone when carrying out a demo. Smartphone Xiaomi Mi 5 with a recording video camera was used with aspect ratio 16:9 and a resolution of 1920x1080, the orientation of the screen: landscape. When recording video for 5-6 seconds without the HDR function (High Dynamic Range which means that camera can see more type of colors, different shades). The video takes about 15 MB. In contrast to the plotting of graphs, we used only 1 reective element.

As a prototype we used a smartphone with the ashlight on and the self-reecting elements on the persons head. A ashlight is needed to ensure that the light falling on the element is reected and better detected on the video.

Detecting people in real time is important for a wide range of applications. In this article [6], a two-stage method was developed to detect people in real time in cluttered and dynamic environments with depth data. A set of possible human peaks is created to ensure the inclusion of all human places. To this end, a new physical radius-depth detector (PRD) is proposed for the rapid detection of candidates for humans. The second stage uses a convolutional neural network (CNN), aimed at extracting the function of the upper body of a person automatically, rather than manual processing, and then based on the CNN function, true human candidates are retained, and false ones are ltered. The results of the experiments are determined by four publicly available data sets, including a data set in low light or even complete darkness. The data show that the proposed method can reliably detect a person in a real-time RGB-D video without accelerating the GPU and gives higher accuracy than the currently compared approaches.

An effective method of analyzing human motion based on segmentation of objects and stereomodes is proposed [7]. First, the object area is segmented and the binary images are scaled. Then, the key points on the object diagram are located, and the key points are matched by normalized color cross-correlation as a measure of similarity. Parallax values of key points are estimated by comparison results, which can be used to analyze the movement of a person. The results show the effectiveness and effectiveness of the method of analysis of human movement.

Simple particle lter by Eiji Ota was used as an algorithm in order to carry out the recognition. The particle lter essentially starts with a bunch of samples (called particles), evolves the state by running each particle through the state equation and re-samples the particles based on the observation that can be seen in order to make the distribution of particles consistent with the observations.

III. IMPLEMENTATION

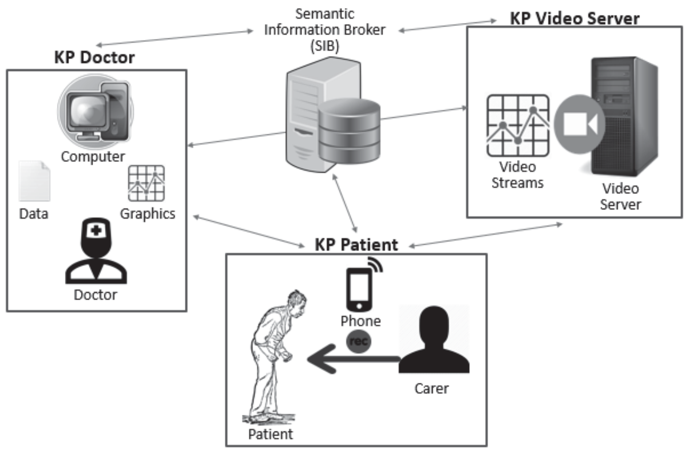

Fig. 2. Medicine architecture

The presented architecture Fig. 2 shows the relationship with which the specific medical prototype was built. Main objects of architecture:

Caregiver for the patient, using the application, turns on the camera and starts to record the patient for a some time on the smartphone. after the end of the recording, he presses the button to stop recording. After that video is sending to the SIB. The patient performs the exercises: get up from the chair, go from one corner to the other (more specifically described in [3]). SIB analyzes the video, depending on the available equipment: on the retro-reflective element (analogy to Biosoft 3D) or on the movement of the head. These two methods have their advantages and disadvantages, it is necessary to use them depending on the situation (the surrounding, the absence of white light on the background, non-availability of reflective element). According to these data, a graph of the movement of the patient's head (with or without a filter) is constructed depending on the time and depending on the frames. Such data is sent to the doctor on the computer and on the video server. The doctor analyzes the findings and leaves his comment about the patient's gait and general health generally. The data is sent again to SIB and to the smartphone of the caregiver for the patient. The video server acts as a data warehouse, storing all the information.

Requirements for computer vision algorithms:

1) Work with low resolution images obtained from video streams with different characteristics;
2) Resistance to image scaling (compression).

Requirements for the subsystems of the software package:

1) Possibility of open access to a local network with video cameras (no more than 10);
2) Finding the identified objects on the image for the purpose of analyzing their movement;
3) Continuous operation of the video server for processing and storing video streams and server for semantic data integration.

Requirements for IoT environment communication resources: For the correct operation of the application, stable operation of the Internet (transmission speed of at least 512 KB/s) is necessary, the Internet itself must be stable (work smoothly), in case of Internet network termination, the request for processing or obtaining the result is automatically re-done. Requirements for the resources of the surrounding equipment: Camcorders that have open access and transmit stream to the Internet. Types of cameras: static, managed IP-cameras, wireless (operating from a network of Wi-Fi). Camcorders should be ”open”. Number of cameras: no more than 10, simultaneous processing of the first image from the camcorder.

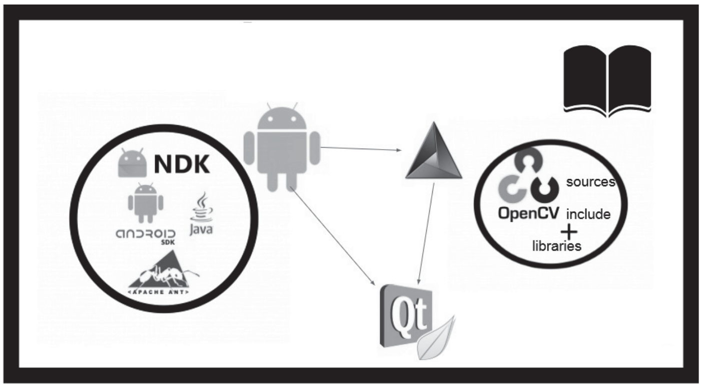

Fig. 3. Application architecture

On Fig. 3 the architecture of the proposed prototype is presented. It is developed on the platform IDE Qt Creator. This prototype is based on the Android platform (versions 5.0 and above). Instead of Android, any other platform supported by the Qt Creator can be used but in this example only Android is shown. Main components for Android are: Android SDK, Android NDK, Java, Apache Ant. Components for CMake are needed to assemble static libraries, suitable for a specific version of Qt Creator IDE and the components for Android. CMake collects source codes, implemented functions and libraries into itself. Android and CMake characteristics are transferred to Qt Creator, then a prototype is built.

In Fig. 4 we can see the work of prototype. We can see the built-in color histogram by colors red, green, blue. The larger value of one of the colors, the higher graphical indicators are. (Than the value of one of the colors is larger, the graphical indicators are higher)

IV. Conclusion

This paper reviewed the possibility of smart interaction of video camera with the environment of smart spaces using

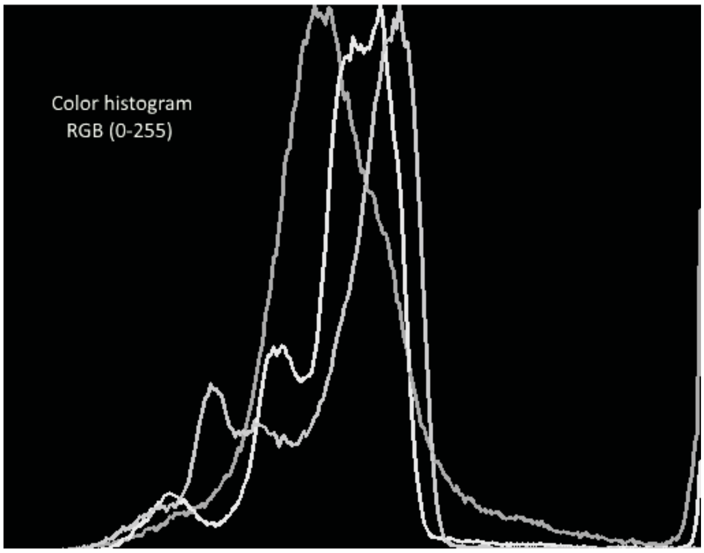

Fig. 4. Color histogram based on color on image

IoT-devices. A prototype was used based on two services: medicine and computer sight, which allows to recognize faces and objects and to build graphics based on camera image recognition.

Acknowledgment

The research was financially supported by the Ministry of Education and Science of Russia within project # 2.5124.2017/8.9 of the basic part of state research assignment for 2017–2019. The reported study was funded from Russian Fund for Basic Research according to research project # 16-07-01289. The results were implemented by the Government Program of Flagship University Development for Petrozavodsk State University in 2017–2021.

References

[1] M. Ajmal, M. Naseer, F. Ahmad, and A. Saleem, “Human motion trajectory analysis based video summarization,” pp. 550–555, 2017.

[2] N. Bazhenov and D. Korzun, “Smartphone-based motion video trackng in patients with parkinson’s disease,” pp. 402–403, 2017.

[3] A. Meigal, K. Prokhorov, N. Bazhenov, L. Gerasimova-Meigal, and D. Korzun, “Towards a personal at-home lab for motion video tracking in patients with Parkinson’s disease,” pp. 231–237, 2017.

[4] D. Korzun, S. Marchenkov, A. Vdovenko, A. Borodulin, and S. Balandin, “Performance evaluation of Smart-M3 applications: A SmartRoom case study,” in *Proc. 18th Conf. of Open Innovations Association FRUCT*, S. Balandin, T. Tyutina, and A. Levina, Eds. ITMO University, Apr. 2016, pp. 138–144.

[5] A. Saha, Y.-W. Lee, Y.-S. Hwang, K. E. Psannis, and B.-G. Kim, “Context-aware block-based motion estimation algorithm for multimedia internet of things (iot) platform,” pp. 163–172, 2017.

[6] J. Zhao, G. Zhang, L. Tian, and Y. Q. Chen, “Real-time human detection with depth camera via a physical radius-depth detector and a cnn descriptor,” pp. 1536–1541, 2017.

[7] M. Ding, G. Sun, N. E. Mastorakis, and X. Zhuang, “An efficient method of human motion analysis on dual-camera stereo vision system,” *ICCAIRO*, vol. abs/1512.03131, 2017.