TALLINNA TEHNIKAÜLIKOOL
MEHAANIKATEADUSKOND

Department of mechatronics

Chair of Mechatronics Systems.

MHK70LT

Igor Yanovskiy

Development of the method for determining the location and orientation of an object in space by using a video sensor.

Master's Thesis

Author applying for

master's sciences of technical

academic degrees

Tallinn
2015

AUTHOR'S DECLARATION

I hereby declare that this thesis is the result of my independent work.

On the basis of materials not previously applied for an academic degree.

All materials used in the work of other authors are provided with corresponding references.

The work was completed .... guidance

".........".....................201….a.

The author

....   signature

The work meets the requirements for a master's work.

".........".....................201….a.

Supervisor

....   signature

Permit to defense

.... curriculum defense superior

".........".....................201… a.

....   signature

TUT Department of Mechatronics

Chair of ....

MASTER'S THESIS SHEET OF TASK'S

Year 2015 semester 2

Student: Igor Yanovskiy 145283   (name, cod)

Curricula: Mechanics

Spetsiality: Mechatronics

Supervisor: Research Scientist, Dmitry Shvarts   (position, name)

Advisers: PhD, Georgiy Zamoruev. (name, position, phone)

MASTER'S THESIS TOPIC:

(in English) Development of the method for determining the location and orientation of an object in space by using a video sensor.

(in Estonian) Loomine määramise meetod asukoha ja orientatsiooni objekti ruumi kasutades video sensor.

Thesis tasks to be completed and the timetable:

Nr	Description of tasks	Timetable
1.	**Decide on the goal of the thesis**	**25.03.2015**
2.	**Calibration of camera in MATLAB**	**03.04.2015**
3.	**Investigate various descriptors**	**20.04.2015**
4.	**Make associating of the coordinates**	**27.04.2015**
5.	**Written part of the work finished**	**14.05.2015**

Solved engineering and economic problems: Calibrated camera using checkboard, and finding of relationship between images in MATLAB.

Additional comments and requirements: …….…………………………...........................

Language: English

Application is filed not later than 12.05.2015

Deadline for submitting the theses 22.05.2015

Student ……………………….. /signature/ ……….................   date ………

Supervisor……………………… /signature/ ………........   date………

Confidentiality requirements and other conditions of the company are formulated as a company official signed letter

Contents

Acknowledgments

In this section I would like to thank the Tallinn Technical University and ITMO University, for the given opportunity to study according to the double-degree program. Also I want to thank departments of mechatronics at both universities and the research supervisor Dmitry Shvarts. And special thanks to my group with which I studied 6 years.

Abbreviations

GPS - Global Positioning System

DGPS - Differential Global Positioning System

INS - Inertial Navigation System

SLAM - Simultaneous Localization And Mapping

SIFT - Scale Invariant Feature Transform

SURF - Speeded Up Robust Features

MSER - Maximally Stable Extremal Regions

GLOH - Gradient Location-Orientation Histogram

Introduction

Computer vision - theory and technology of creating machines that can produce detection, tracking and classification of objects.

As a scientific discipline, computer vision is concerned with the theory and technology of artificial systems that obtain information from images. The video data may be represented by a set of forms, such as a video sequence images from different cameras or three-dimensional data, such as camera, medical devices or Kinect scanner.

As a technological discipline, computer vision seeks to apply the theories and models of computer vision to create computer vision systems. Examples of such systems are:

1. Process control systems (industrial robots, autonomous vehicles)
2. Systems of video surveillance
3. Organizing information systems (for example for indexing databases of images)
4. System modeling of objects or the environment (analysis of medical images, topographical modeling)
5. System interaction (for example, an input device for a system of man-machine interaction)
6. Augmented Reality
7. Computational picture, for example mobile cameras

Computer vision can also be described as a complement (but not necessarily the opposite) of biological vision. In biology studied human visual perception and different animals, thereby creating models of such systems in terms of physiological processes. Computer vision, on the other hand, describes studies and computer vision systems, which are made in hardware or software. Interdisciplinary exchange between biological and computer vision has been very productive for both scientific fields.

Subsections computer vision includes playing action event detection, tracking, pattern recognition, image restoration and others.

In this paper we will have developed a method for finding the position and orientation of the object using a video sensor. It is necessary to create 3D scenery on stage.

The Scientific Objectives of the Thesis

The aim of this work is to develop the method to determinate location and orientation of an object in the space using video sensor.

In order to successfully solve the problem, it is necessary to solve a number of important tasks:

- calibrate camera
- determine the relationship between two images
- find the position and orientation of the camera after a moving it from one position to another

It could be done using mechanical sensor, but we will do the same with one exception: using computer vision and mechanical knowledge. This method can be used in performances without decorations.

Referaat

See lõputöö on kirjutatud Tallinna Tehnikaülikoolis Tallinnas. Töö eesmärgiks on video sensori abil objekti asukoha ja orientatsiooni määramise metoodika väljatöötamine. Töö käigus said lahendatud järgmised ülesanded: kaamera kalibreerimine, kujutiste vahelise seose leidmine, kujutise ruumilise asukoha määramine.

Esimeses peatükis räägitakse masinnägemisest üldiselt. Kuigi masinnägemise alased uuringud algasid 1970ndate aastate lõpus, pole paljusid teemaga seotud probleeme siiani lahendatud.

Teises peatükis tutvustatakse olemasolevaid objekti asukoha ja orientatsiooni määramise meetodeid kus kasutatakse erinevaid andureid ja süsteeme. Põhimõtteliselt põhinevad kõik need meetodid mehaanilisel mõõtmisel ja anduritel mis võimaldavad saada vajalikku informatsiooni. Ainult üksikud meetodid kasutavad vajaliku informatsiooni saamiseks kontaktivaba mõõtmist.

Kolmandas peatükis on mõningate deskriptorite ja detektorite kirjeldused. Esitatud deskriptorid ja detektorid võimaldavad leida samasuguseid punkte erinevatelt, samade elementidega piltidelt. Sarnasuste leidmiseks piltidelt kasutati Harrise meetodit.

Neljandas peatükis kirjeldatakse kaamera kalibreerimist kaamera sisemiste parameetrite kindlakstegemiseks. Kalibreerimiseks on kasutatud Matlab Camera Calibration Toolbox'i. Kalibratsiooni kujutise e. "scene" jaoks kasutati malelaua kujutist.

Viiendas peatükis kirjeldatakse kujutistelt sarnasuste leidmise ja muundemaatriksi leidmise protsessi. Samuti kirjeldatakse ka seose leidmist koordinaatide ja kujutise punkti vahel.

Töö tulemusena on Matlab programmi abil välja töötatud meetod objekti asukoha määramiseks liikuva kaameraga. Meetodit võib kasutada teatrietenduste puhul kus pole dekoratsioone. Idee seisneb selles, et laval on ainult objekt, dekoratsioonid aga liiguvad või vahetuvad automaatselt sõltuvalt kaamera või obekti asukohast või orientatsioonist.

Abstract.

This thesis was carried out at the Tallinn Technical University in Tallinn. The purpose of this work is development of the method for determining the location and orientation of an object in space by using a video sensor. In the course of work the following tasks were solved: calibration of the camera; finding of the relationship between images; finding of dependence of coordinates between the camera and a point in the space.

In chapter 1 is narrated about computer vision in general. Although studies based on computer vision began in late 1970, but there are still many problems and are not solved.

The second chapter describes the existing methods and systems for determining the location and orientation of the object by using various sensors, systems and methods. Basically, all these methods are based on the dimensions using mechanical sensors to recognize the information. And only a few methods provide the necessary information on the basis of non-contact measurement.

In the third chapter, you can see the description of certain descriptors and detectors. Presented tags and detectors allow us to find the same point on different images with the same elements. As a method of finding matches the images was chosen Harris detector.

In the fourth chapter there is a camera calibration to determine the internal parameters of the camera. Calibration is carried out using Matlab Camera Calibration Toolbox. As a "scene" was taken chessboard.

The fifth chapter describes how to find the going match the images and find the translation matrix. As well as finding the dependence of coordinates between the camera and the point on this "scene".

In the end, it turns out to develop a method using the program Matlab. This method is developed for determining the location of an object while moving the camera. It can be used in the theater without scenery. The idea is that the stage would only action object and scenery will automatically move or vary depending on the location and orientation of the camera or object.

1. Computer vision

1.1 Brief History

Computer vision — the theory and technology of creation of cars which can make detection, tracking and classification of objects.

Computer vision can be described as a young, diverse and dynamic. And although there are earlier works, we can say that only the end of 1970 began an intensive study of the problem, when computers could manage the processing of large sets of data, such as images. However, these studies usually started from other areas, and therefore, there is no standard definition of the problem of computer vision. Also, and this is even more important, there is no standard formulation of how the problem should be solved computer vision. Instead, there are a lot of methods to solve various problems strictly defined computer vision, where the methods are often dependent on tasks and seldom can be generalized to a wide range of applications. Many of the techniques and applications are still in the stage of basic research, but an increasing number of methods is used in commercial products, where they often form part of a larger system that can solve complex problems (for example, in the field of medical imaging and measurement and quality control manufacturing processes). In most practical applications of computer vision pre-programmed computers for the solution of certain problems, but the methods, knowledge-based, are becoming more common.

An important part in artificial intelligence takes automatic planning or decision-making systems that can perform mechanical actions such as moving a robot through some medium. This type of treatment usually needs input data provided by the system of computer vision, image sensor, and acting as providers of high-level information about the environment and the robot. Other areas that are sometimes described as belonging to artificial intelligence and are used with respect to computer vision is pattern recognition and learning techniques. As a result, computer vision is sometimes seen as part of the field of artificial intelligence and computer science in general.

Another area related to computer vision is signal processing. Many of the treatment methods of one-dimensional signals, typically temporal signals can be naturally extended to the processing of two-dimensional or multidimensional signals in computer vision. However, because of the peculiar nature of the image, there are many techniques

developed in the field of computer vision and have no analogues in the treatment of one-dimensional signals. A particular feature of these methods is their linearity, that together with multidimensional signal makes a corresponding subdomain of signal processing part areas of computer vision.

In addition to these approaches to computer vision, many of the test questions can be studied from a purely mathematical point of view. For example, many techniques are based on statistics, the optimization methods or geometry. Finally, the extensive work carried out in the practical application of computer vision - of the existing methods may be implemented in hardware or software, and they can be modified in order to achieve high-speed operation without significant increase in resources consumed.

1.2 Examples of applications of computer vision

One of the most important applications is medical imaging. This area is characterized by obtaining information from the image data for medical diagnosis to patients. In most cases, the video data obtained by microscopy, X-ray, angiography, ultrasound and tomography. Examples of information that can be obtained from such video data is the detection of tumors, atherosclerosis or other malignant changes. Also, an example would be the measurement of dimensions, and so blood flow. D. This application area also contributes to medical research, providing new information, for example, on the structure of the brain or the quality of medical treatment.

Another application area of computer vision is an industry. Here, information is obtained for the purpose of supporting the production process. An example is the control of quality when the final product or parts are automatically checked for defects. Another example is the measurement of the position and orientation of the parts that are raised by a robot arm.

Military applications is probably the biggest area of computer vision. Obvious examples are the detection of enemy soldiers and vehicles and missile control. The most advanced control system with missiles sent a rocket into the specified area instead of a specific purpose, and the purpose of the selection is made when the missile reaches the target area, based on the received video data. Modern military concept such as "combat awareness" means that a variety of sensors, including image sensors provide a large set of information on the field of battle, which can be used to make strategic decisions. In this case, the automatic data processing is used to reduce the complexity and increase the reliability of the information received.

One of the new applications are autonomous vehicles, including submarine, terrestrial (robots, machines), air. The level of autonomy ranges from fully autonomous (unmanned) vehicles to which systems based on computer vision, supported by the driver or a pilot in various situations. Fully autonomous vehicles using computer vision for navigation, ie to obtain information about his whereabouts, to create a map of the environment, to detect obstacles. They can also be used for various tasks, e.g., for the detection of forest fires. Examples of such systems may be a system warning of obstacles on the machines and systems of autonomous landing. Some car manufacturers have demonstrated an autonomous driving system, but the technology still has not reached a level where it can be run in mass production.

Other applications include:

- Support for the creation of visual effects for film and television
- Watching

1.3 Typical tasks of computer vision

Each of the computer vision applications discussed above, related to a number of problems; more or less well-defined measurement or processing problems can be solved by using a plurality of methods. Some examples of typical computer vision tasks are presented below.

Recognition

The classical problem in computer vision, image processing and machine vision definition video contain any object a certain characteristic, feature or activity. This problem can be solved easily and reliably man, but has not yet been satisfactorily solved in computer vision in general: random objects in random situations.

Existing methods for solving this problem are effective only for the individual objects, such as simple geometric objects (eg, polyhedrons), human faces, printed or hand-written characters, vehicles, and only in certain circumstances, it is usually a certain lighting, background and position of the object relative to the camera.

In the literature, different set of problems of recognition:

- Detection: one or more predetermined or learned objects or classes of objects can be detected, usually with their position in the two-dimensional image or three-dimensional position in the scene.
- Identification: recognize an individual instance of an object. Examples: identification of specific human faces or fingerprints or car.
- Detection: video checked for the presence of certain conditions. For example, the possibility of incorrect detection of cells or tissues in medical images. Detection based on relatively simple and fast calculations are sometimes used to find small areas in the analyzed image, which is then analyzed using techniques more demanding, for a correct interpretation.

There are a number of specialized tasks based on the recognition, for example:

- Image search content: finding all images in a large set of images that have a certain content. The content can be defined in various ways, for example in terms of similarity with the specific image (find me all the pictures related to this image), or in terms of high-level search criteria entered as text data (I get all the pictures, in which many houses are made in winter, and in which no cars).
- An evaluation of: determining the position or orientation of the object relative to a particular camera. An example of application of this technique may be to promote the robotic arm in retrieving objects from the conveyor belt to the assembly line.
- Optical Character Recognition: character recognition on images printed or handwritten text, usually to convert into a text format that is most convenient for editing or indexing (for example, ASCII).

Motion

Several problems related to motion estimation in which a sequence of images (video) processed for finding a bound rate of each point in the image or 3D scene. Examples of such tasks are:

- Determining the three-dimensional movement of the camera
- Tracking, that is following the movements of the object (for example, machines or people)

Recovering stage

Given two or more images of the scene, or video data. Restoring the scene has the task to recreate a three-dimensional model of the scene. In the simplest case, the model can be a set of three-dimensional space of points. More sophisticated methods of reproducing the full three-dimensional model.

Image restoration

The task of restoring the image is the removal of noise (sensor noise, motion blur, and so on. D.). The simplest approach to this problem are the different types of filters, such as filters or lower midrange. More sophisticated methods use representations that should look like these or other areas of the image, and on the basis of their variation.

A higher level of noise removal is achieved in the initial analysis for the presence of video data of different structures, such as lines or borders, and then filtering process control based on these data.

1.4 Computer Vision Systems

Implementation of computer vision systems is strongly dependent on its use, and the hardware platform performance requirements. Some systems are self-contained and solve specific problems of detection and measurement, whereas other systems make up the sub-systems of larger systems that may already contain control subsystem mechanical manipulators (robots), Information Database (search for similar images), human-machine interfaces and (computer games), etc. However, there are features that are typical for many computer vision systems.

- Obtaining image: digital images are obtained from one or more image sensors, and that in addition to various types of light-sensitive cameras include distance sensors, radar, cameras, ultrasonic, etc.. G., Depending on the sensor type, the data can be obtained in the usual 2D image, 3D image or the sequence of images. Usually correspond to pixel values of the light intensity in one or more spectral bands (color or grayscale), but may be associated with different physical measurements, such as depth, absorption or reflection of sound waves or electromagnetic, or nuclear magnetic resonance.
- Pretreatment: before the computer vision techniques can be applied to image data in order to extract the certain amount of information necessary to process the video

data so that they satisfy certain conditions, depending on the method used. Examples are:

- o Re-sampling in order to ensure that the image coordinate system is correct
- o Noise removal in order to remove distortions introduced sensor
- o Contrast range improving in order that the necessary information could be found
- o Scaling to better distinguish between the structures in the image

- Allocation details: details of the image different levels of complexity are extracted from the video data. Typical examples of such products are:
  - o Lines, borders and edges
  - o Localized interest points such as corners, drops or dots: more complex parts can relate to the structure, form and movement.

- Detection / Segmentation: at a certain stage of processing, the decision on what points or portions of the image are important for further processing. Examples are:
  - o Selecting a specific set of points of interest
  - o Segmentation of one or more areas of the image that contain distinctive item

- The high-level processing: in this step the input data is typically a small set of data, such as a set of points or a portion of the image, which is presumed to be a particular object. Examples are:
  - o Check that the data satisfy the conditions depending on the method and application
  - o Evaluation of the characteristic parameters, such as position or size of an object
  - o The classification of the detected object into different categories

1.5 Conclusion

This chapter presents the history of computer vision and review of progress since the early work on the subject. There is also considered the problems and examples of computer vision. In next chapter we will review existing methods of determination orientation and location some objects.

2. Review of existing methods and systems

2.1 Introduction

In recent years there have been developed a variety of methods and the invention allows to determine the location and orientation of an object in space. Many tasks have been solved. Some methods used directly in any mechanical sensors. These methods mostly use accelerometers and gyroscopes. There are very few methods of using non-contact measurement.

2.2 Accelerometer

Accelerometer (lat. Accelero - faster and ancient Greek. Μετρέω "measure") - a device that measures the projection of the apparent acceleration (the difference between the true acceleration of the object and the gravitational acceleration). As a rule, the accelerometer is sensitive mass designated in an elastic suspension. Deviation from the weight of its original position if the apparent acceleration carries information about the magnitude of this acceleration (Fig.1).
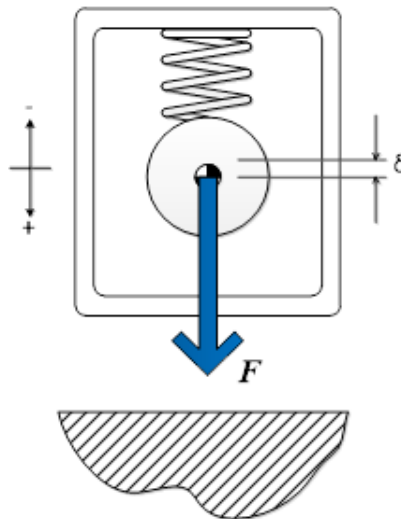

Fig. 1 Accelerometer

On a design accelerometers are divided into one-component, two-component, three-component. Accordingly, they can measure the acceleration along one, two or three axes.

Some accelerometers also have the built-in systems of collecting and data processing. It allows to create complete systems for measurement of acceleration and vibration with all necessary elements.

2.2.1 Parameters

The main parameters of the accelerometer are:

- Large-scale coefficient — proportionality coefficient between the measured seeming acceleration and an output signal (an electric signal, frequency of fluctuations (for the string accelerometer) or a digital code).
- Threshold sensitivity (permission) — the size of the minimum change of the seeming acceleration which the device is capable to define.
- Shift of zero — instrument reading at the zero seeming acceleration.
- Random walk — a mean square deviation from shift of zero.
- Nonlinearity — changes of dependence between an output signal and the seeming acceleration at change of the seeming acceleration.

2.3 Gyroscope

Gyroscope (from ancient Greek. Γῦρος - Circle + σκοπέω - look) - a device that can respond to changes in the angles of orientation of the body in which it is established with respect to an inertial reference system. The simplest example of a gyroscope - yul (top).(Fig.2.)
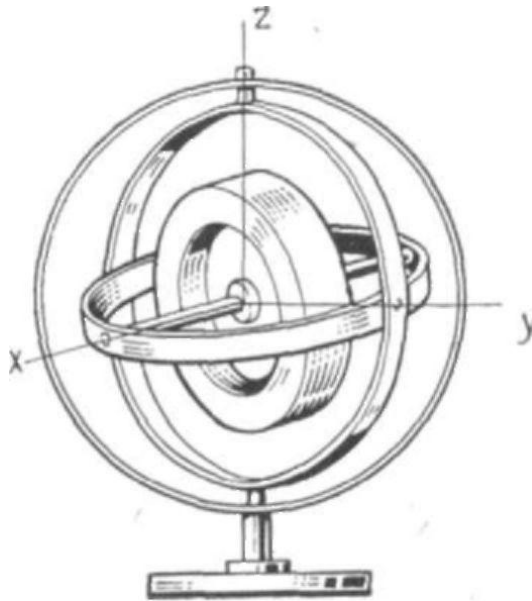
Fig. 2 Gyroscope

2.3.1 Classification

The main types of gyros on the number of degrees of freedom:

- two-degree,
- the three-degree.

The main two types of gyros on the principle of action:

- mechanical gyroscopes,
- optic gyroscopes.

Also researches on creation of the nuclear gyroscopes using a nuclear magnetic resonance for tracing of change of a spin of atomic kernels are conducted [1].

Mechanical gyroscopes

Among mechanical gyroscopes the rotor gyroscope — quickly rotating solid body (rotor) which spin axis can freely change orientation in space is selected. Thus rotational speed of a gyroscope considerably exceeds the speed of turn of an axis of its rotation. The main property of such gyroscope — ability to save in space the invariable direction of a spin axis in the absence of impact of the moments of external forces on it and effectively to

resist action of the external moments of forces. This property substantially is defined by value of angular speed of own rotation of a gyroscope.

Properties of a three-degree rotor gyroscope

In case of influence of the moment of external force round an axis, a perpendicular spin axis of a rotor, the gyroscope starts being turned round an axis of precession which is perpendicular to the moment of external forces(Fig.3).
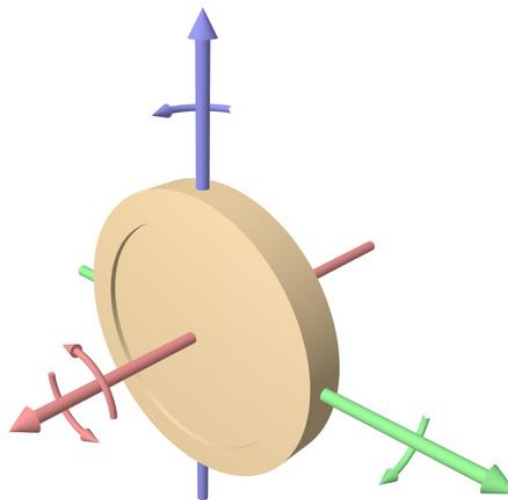

Fig. 3 Precession of a mechanical gyroscope.

Optical gyroscopes

Gyroscopes, passive optical gyroscopes, fiber optical and integral and optical share on laser (the active optical). The principle of action is based on Sanyak's effect opened in 1913 [2].

Theoretically he has a talk with the help of the special relativity theory. According to the special relativity theory light velocity is constant in any inertial frame of reference [3]. While in noninertial system it can be differ [4]. When sending a ray of light in a rotation sense of the instrument and against a rotation sense the time difference of arrival of rays (determined by the interferometer) allows to find a difference of optical ways of rays in inertial frame of reference, and, therefore, value of angular turn of the instrument during passing of a ray. Value of effect is directly proportional to the angular speed of rotation of the interferometer and the area enveloped by a propagation path of light waves in the interferometer.

Application in technic.

Properties gyroscope used in devices - gyros, most of which is a rapidly rotating rotor, which has several degrees of freedom (axes rotatably).

The most commonly used gyroscopes placed in the gimbal. These gyros have three degrees of freedom, that is, it can make 3 independent rotation about the axis AA ', BB' and CC ', intersecting at the center of suspension O, which remains in relation to the base A fixed.

Gyroscopes, whose center of gravity coincides with the center of the suspension O, called astatic otherwise - static gyroscopes.

Special gyro motors are applied to support of rotation of a rotor of a gyroscope with a high speed.

For control of a gyroscope and removals of information from it are used sensors of an angle and torque sensors.

Gyroscopes are used in the form of components as in systems of navigation (a gyrohorizon, a gyro-compass, inertial navigation, etc.), and in systems of orientation and stabilizing of spacecrafts. When using in a gyrovertical of the indication of a gyroscope shall be adjusted by the accelerometer (pendulum) as because of daily rotation of the earth and leaving of a gyroscope, there is a deviation from truth of a vertical. Besides, in mechanical gyroscopes offset of its center of masses which is equivalent to direct impact of a pendulum on a gyroscope can be used [5].

2.4 Echolocation and hydrolocation

Echolocation (echo and lat. locatio — "position") — the way in which the position of an object is determined by the delay time of the reflected wave returns. If the waves are sound, this sonar, if the radio - radar.

Hydrolocator or sonar (an abbreviation of SOund Navigation And Ranging) - a means of sound detection of underwater objects using acoustic emission. In the UK, until 1948 he used the name "asdik» (ASDIC, an abbreviation of the Allied Submarine Detection Investigation Committee).

2.4.1   Basic information

According to the principle of action sonars are:

- Passive - allowing to determine the location of the object at the position of the underwater sound signals emitted by the object (noise direction finding).
- Active - using reflected or scattered underwater object signal emitted by his side sonar.

Active sonar "Asdik" in its original primitive form was invented at the end of the First World War. The basic principle of his actions has remained unchanged to this day. However, over the years greatly increased the effectiveness of sonar, widened the scope of its use, as well as increased the number of classes of ships, with which it could be used for search and attack enemy submarines.

The basis is a transceiver that sends sound pulses in the desired direction, and receives the reflected pulses, if the parcel, meeting on its way any object that will affect him. These assumptions and reflected signals after converting sound very similar to how to pronounce the word "ping". Therefore, it became known as ping set, the work on it was called pinging, a specialist officer for anti-submarine warfare - pinger.

Turn the transceiver like a searchlight, you can determine the compass direction in which the ping is sent, and hence the direction of the object, which is reflected ping. Noticing the time between sending the pulse and the reception of the reflected signal, you can determine the distance to the object detected.

2.4.2   Radiolocation

Radiolocation — the area of science and technique integrating methods and means of a location (detection and measurement of coordinates) and determination of properties of different objects by means of radio waves [6] [7] [8]. To relatives and partly the superimposed term is the radio navigation, however radio navigations more active role is played by object which coordinates are measured, most often this determination of own coordinates. The main technical adaptation of radiolocation — radar station(Fig.4.).
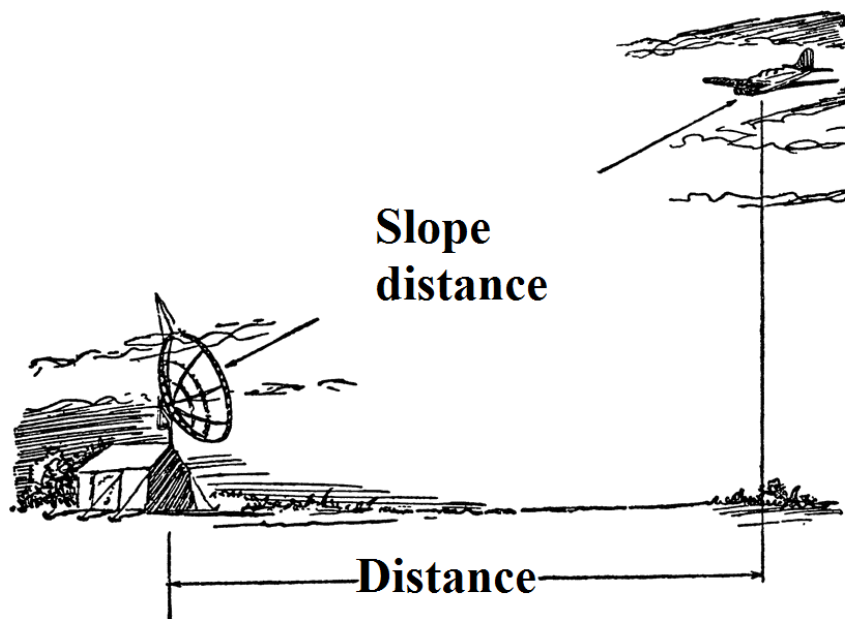
Fig. 4 Example of radiolocation

Radiolocation as science is based on use of a row of the physical laws connected to distribution and dispersion of electromagnetic waves (EMW). The property of electromagnetic waves, major for radiolocation, is their dispersion when falling on objects. It allows, accepting the waves dispersed by object and measuring their parameters, to judge existence and object properties. Generally the object disperses waves extensively, including aside, reverse to the direction of arrival of incident wave. Thus, irradiating object, it is possible to accept the reflected mode in the same place, the radiation wave from where extends.

The law on linearity of distribution of electromagnetic waves in a homogeneous environment is used for determination of angular data of object by measurement of the direction of arrival of a wave as vector, perpendicular to phase wave front.

The persistence of speed of distribution of electromagnetic waves allows to define range to object by measurement of a time delay of a wave in case of distribution it from object to the radar.

2.5 GPS

2.5.1 Introduction

GPS (Global Positioning System) – a satellite navigation system for the measurement of distance, time and determines mestopolozhenie in the global coordinate system WGS

84. Allows you to anywhere in the earth (excluding polar regions), in almost any weather, as well as in near-Earth space to locate and speed of the objects. The system is designed, implemented and operated by the US Department of Defense, and the currently available for use for civilian purposes – only the navigator or other device (eg smartphone) with a GPS-receiver(Fig.5).
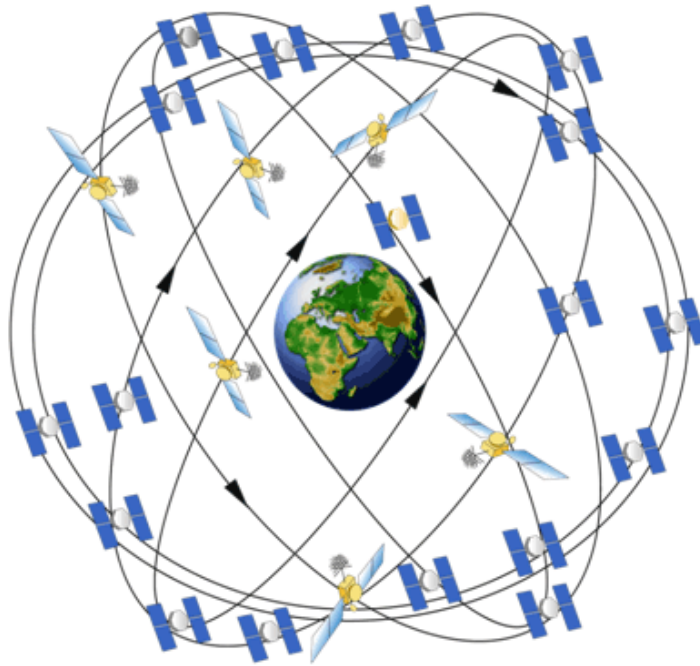

Fig. 5 GPS

The basic principle of using the system – positioning by measuring the reception timing of the synchronized signal from the navigation satellite antenna consumer. To determine the three-dimensional coordinates of the GPS-receiver need to have four equations: "Distance is the product of the speed of light on the difference between the time of signal reception by the consumer and the date of its synchrotron radiation from satellites»: $|x - a_j| = c(t_j - \tau)$. Here: a_j - j th location satellite, t_j - time reception of j th satellite's clock customer, τ - an unknown time synchrotron radiation signal all satellites bounds consumer c - the speed of light, x - an unknown three-dimensional position of the user.

GPS consists of three segments: space, control and user [9]. GPS satellites broadcast signals from space, and all the GPS receiver uses this signal to calculate its position in the space of three coordinates in real time.

The space segment consists of 32 satellites orbiting at a medium Earth orbit. The control segment is the main control station and several additional stations [10] and ground antennas and monitoring stations, some of these resources are shared with other projects.

The user segment is represented by the receivers GPS, under the authority of state institutions, and hundreds of millions of devices owned by normal users.

2.5.2 Applications of GPS

Despite the fact that initially design GPS has been directed towards military purposes, today widely used in the GPS civilian purposes. GPS-receivers are sold in many stores selling electronics, they are inserted in mobile phones, smart phones, electronic wrist watches, PDAs and onbordery. Consumers are also offered a variety of devices and software products that allow to see your location on an electronic map; having the ability to run routes, taking into account road signs, permitted turns and even traffic jams; look on the map specific homes and streets, attractions, cafes, hospitals, gas stations and other infrastructure.

- Geodesy: by means of GPS exact coordinates of points and border of the land plots are defined.
- Cartography: GPS is used in civil and military cartography.
- Navigation: with application of GPS both sea, and road navigation is carried out.
- Satellite monitoring of transport: by means of GPS monitoring of situation, speed of cars, control of their movement is conducted.
- Cellular communication: the first mobile phones with GPS appeared in the nineties. In some countries, for example it is used by the USA for expeditious definition of location of the person calling 911. In Russia in 2010 implementation of the similar project — ERA-GLONASS is begun.
- Tectonics, Tectonics of plates: by means of GPS are observed movements and fluctuations of plates [11].
- Local Activities: There are different games, where applicable GPS, such as caching and other.
- Geotagging: information, for example photos, "become attached" to coordinates thanks to the built-in or external GPS receivers.

Offers on integration of Iridium and GPS systems expressed [12].

2.5.3   DGPS

DGPS (differential global positioning system) — the system of increase of accuracy of signals of GPS consisting in the account and measurement of a difference between known pseudo-ranges to satellites and the actual code pseudo-ranges [13]. Thus both atmospheric distortions, and ephemeris mistakes are compensated. Main sources of signals of DGPS:

- radio navigational beacons;
- satellites in a geostationary orbit.

Radio beacons transmit DGPS signals at average frequencies in the range of 283,5-325 kHz. As radio signals are reflected in these frequencies from a terrestrial surface, the hilly and mountain district usually doesn't influence reception of a signal. However signals don't reach the deep canyons located far from radio beacons.

2.6 INS

2.6.1   Introduction

INS (Inertial navigation system) — a navigation method (determination of coordinates and parameters of the movement of various objects — vessels, planes, rockets, etc.) and managements of their movement, based on properties of inertia of bodies, being autonomous, i.e. not demanding existence of external reference points or the signals arriving from the outside. Nonautonomous methods of the solution of problems of navigation are based on use of external reference points or signals (for example, stars, beacons, radio signals, etc.). These methods in principle are rather simple, but in some cases can't be carried out due to the lack of visibility or existence of hindrances for radio signals, etc [14]. Need of creation of independent navigation systems was the reason of emergence of inertial navigation (Fig.6).
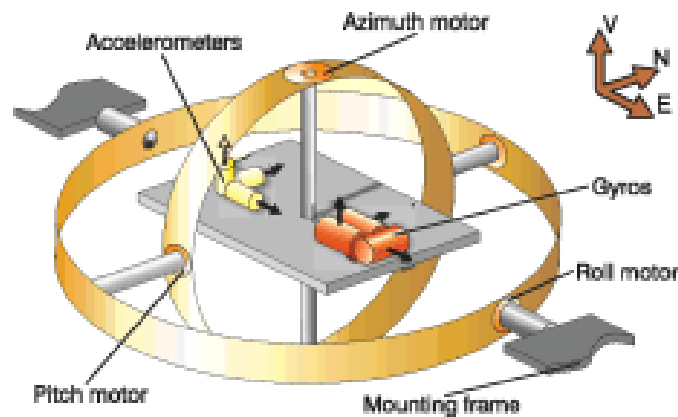
Fig. 6 Inertial navigation system

2.6.2 Operating principle

The essence of inertial navigation consists in definition of acceleration of object and its angular speeds by means of the devices and devices installed on moving object, and according to these data — location (coordinates) of this object, its course, speed, the passable way, etc., and also in determination of the parameters necessary for stabilization of object and automatic control of its movement. It is carried out with the help:

- sensors of linear acceleration (accelerometers);
- the gyroscopic devices reproducing a reference system on object (for example, by means of the gyrostabilized platform) and the angles of rotation and an object inclination used for its stabilization and traffic control allowing to define.
- computers which on accelerations (by their integration) find the speed of object, its coordinates, etc. movement parameters;

Advantages of methods of inertial navigation consist in autonomy, noise immunity and possibility of full automation of all processes of navigation. Thanks to it methods of inertial navigation are widely used more and more at the solution of problems of navigation of surface courts, submarines, planes, spacecrafts and other moving objects.

2.7 Odometry

Odometry (from the Greek words hodos ("movement", "travel") and metron ("measure", "to measure")) — use of the drives given about the movement, for a movement assessment.

Odometry is applied:

- in robots (for calculation of the passable way)
- machines with numerical program control (for tracking of position of the tool),
- printers (for tracking of provision of the printing head),
- control units, for example — computer mice (for movement tracking)

The standard scheme of an odometry of the robot, the use of encoders, reading the angle of rotation of wheels (Fig.7).
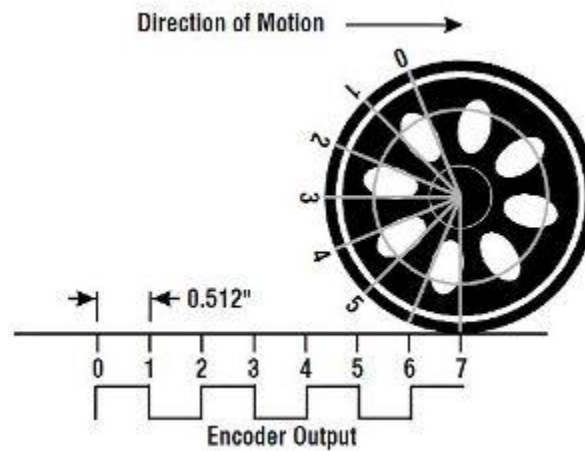

Fig. 7 Standard scheme of an odometry

Using different types of encoders, it is possible to achieve tracking not only a corner, but also the direction of rotation (for example — using Gray's codes).

2.8 SLAM

SLAM - Simultaneous Localization And Mapping - Method of simultaneous navigation and map the - the method used robots and autonomous vehicles for the map in unknown space, or to update the map to known space while controlling the current location and distance traveled.

The method of simultaneous mapping and navigation (SLAM) - a concept that connects two independent processes in a continuous cycle of consecutive calculations, in which the results of the process involved in the computation of another process.

Mapping - is the problem of integrating the information gathered from the sensors of the robot. In this process, the robot as it responds to the question: "What is the world?"

The main aspects in the construction of maps are presentation of data about the environment and the interpretation of sensor data.

Rather, the localization - a problem locating the robot on the map. At the same time, the robot as it answers the question "Where am I?"Localization can be divided into two types - local and global.

Local localization allows you to track the location of the robot on the map when its initial location is known, and the global localization - is positioning the robot in a strange place (for example, in the abduction of the robot).

Thus the method of simultaneous mapping and navigation (SLAM) - it is a problem of mapping and localization of the robot in this map. In practice, these two problems can not be solved independently. Before the robot will be able to answer the question of how the environment looks like (based on a series of observations), it needs to know where the observations were made. At the same time, it is difficult to estimate the current position of the robot without the card.

Moreover, the complexity of calculating the current position and map the confined to low precision devices involved in the process of calculating the current location.

2.9 Chroma key

Chroma key — technology of combination of two and more images or shots in one composition, the color back projection which is used on television and in modern digital technology of film production. During shootings the object is located on a monophonic color background. At combination in an object shot with a background during record of a scene or at installation instead of a background it is possible to place other image. Also in everyday life chroma key call the color back screen against which remove (Fig.8.).

Fig. 8 Example of chroma key

The most widespread flowers which are used at projection are green and blue (blue) because such colors don't meet in tones of human skin. However any color, including white and black can be practically used. The most popular color of a background for the combined shootings in film production — green (that gave the name to the Green screen technology), is more often applied to television programs a blue background ("Blue screen") though color of the screen depends on the creative task set by the director and characteristics of the equipment on which projection is made.

The only minus chroma key is that if on clothes of the person there is a color similar to color of a background, people starts "appearing through" ("to shine"). Therefore at selection of a suit of the actor or TV host avoid the flowers coinciding with a background. In a cinema before emergence of television technologies and digital ways of film production the ways of combination of an actor's scene with a background known as the "back projection", "frontproyektion" and "the wandering mask" giving similar with hromaky graphic effect were applied.

2.10   Motion capture

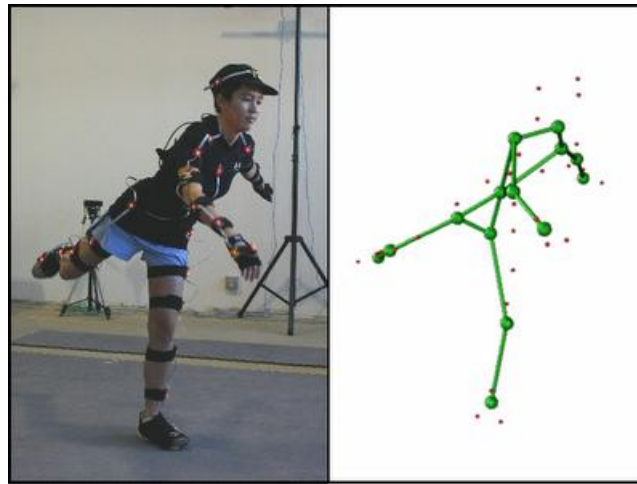Motion capture — a method of animation of characters and objects (Fig.9).

Fig. 9 Example of motion capture

There are two main types of motion capture systems:

- Marker motion capture system where the special equipment is used. The suit with sensors is put on the person, he makes the movements demanded according to the scenario, strikes agreed poses, imitates actions; data from sensors are fixed by cameras and come to the computer where are reduced in the uniform three-dimensional model which is precisely reproducing the movements of the actor on the basis of which later (or in real time) animation of the character is created. Also by this method the actor's mimicry is reproduced (in this case on his face the markers allowing to fix the main mimic movements settle down).

- Markerless technology that does not require special sensors or special costume. Markerless technology is based on the technology of computer vision and pattern recognition. The actor may appear in casual clothes, which greatly speeds up the preparation for the shooting, and allows you to capture complex movements (fighting, falling, jumping, and so on.) Without the risk of damage to the sensors or markers. Several implementable markerless systems have been developed in recent years [15] [16], although studies of this technology have been held for a long time [17]. Today, there are software "desktop" class markerless motion capture. In the present case does not require special equipment, special lighting and space [18]. Shooting is performed using a conventional camera (or webcam) and a personal computer.

Today there are a large number of marker motion capture systems. The difference between them lies in the principle of the transfer of movement [18].

29

- Optical passive. In the suit, included with this system, sensors are attached markers, which are called passive, because they reflect only send them the light, but they do not shine. In such systems, the light (infrared) markers is sent from the camera's strobe high and reflected from the marker gets back into the camera lens, telling it the position of the marker:

    Minus of optical passive systems consists in duration of placement of markers on the actor. Also sometimes at the fast movement or a close arrangement of markers to each other the system can confuse them (as this technology doesn't provide identification of each marker separately).

- The optical active are called so because instead of reflecting markers which fasten to the actor's suit, in them light-emitting diodes with the integrated processors and radio synchronization are used. To each light-emitting diode ID (identifier) that allows system not to confuse markers with each other is appointed, and also to learn them after they were blocked and again appeared in sight of cameras. In all the rest the principle of work of such systems is similar to passive systems.

    Minuses of active systems:
    o No possibility of motion capture and facial expressions
    o Additional controller fixed to the actor and connected to the marker LEDs, binds its traffic
    o Fragility and relatively high cost marker LEDs
- Magnetic systems in which markers are magnets and the chamber - a receiver, the system calculates its position on the distortion of the magnetic flux.

    Minuses of magnetic systems:
    o Magnetic systems are subject to magnetic and electrical interference from metal objects and environments (electrical facilities, office equipment, fixtures in plates of the structure)
    o Changeable sensor sensitivity depending on their position in the work area
    o Less in comparison with optical systems work area
    o No possibility of motion capture and facial expressions
    o Additional controller attached to a cast and is connected to the magnetic markers, or even a bunch of wires stretching from the actor to the computer.
    o The high cost of magnetic markers

- Mechanical systems directly follow the folds of the joints, for this actor to wear special mocap-mechanical skeleton that repeats after him all the movements. In the computer while data is being transferred on the corners bend all the joints.

  Minuses of mechanical systems:
  - Mosar-skelet, with an additional controller, attached to the actors and sensors connected to the fold, and in some cases, the wires extending from the skeleton, severely limit the movement of the actor.
  - The absence of the possibility of capture:
    - Movements and facial expressions
    - Movements of close interaction between two or more actors (wrestling, dancing with support, and etc.)
    - Movements on a floor — somersaults, falling etc.
  - The risk of damage by careless use of mechanics
- Gyroscopic / inertial systems for collecting traffic information using miniature gyroscopes and inertial sensors placed on the body of the actor - as well as markers or magnets in other mocap-systems. The data from the gyroscopes and sensors are transmitted to a computer, where they are processed and recorded. The system determines not only the position of the sensor, but also the angle of it.

  Minuses of gyroscopic / inertial systems:
  - No possibility of motion capture and facial expressions
  - Additional controller attached to a cast and is connected to the magnetic markers, or even a bunch of wires stretching from the actor to the computer.
  - The high cost of gyroscopes and inertial sensors
  - To determine the position of the actor in the space need more mini-system (optical or magnetic)

2.11   Conclusion

In second chapter you have seen review of some existing methods, sensors and systems which helps to determinate location and orientation of objects. Here presents a large numbers of different solutions. But most of them use mechanical sensors. In our solution we will review method based on computer vision. To solve our problem we will use some descriptors and detectors. But at first we will consider various descriptors in the following chapter.

3. Review of descriptors

3.1 Introduction

The handle (from the Latin. Descriptor - describing) - lexical unit information retrieval language used to describe the semantic content of the basic instrument or formulation of the query when searching for a document (information) in a retrieval system. The handle is assigned a unique group of keywords natural language taken from the text relating to a particular field of knowledge.

Image Descriptor defines the actual location and razmeryposleduyuschego image within the space defined by the Screen Descriptor. As determined by the flag indicating the presence of local color lookup table and sequencing highlight pixels. This section provides an overview of some existing descriptors.

The input data descriptor is an image and a set of singular points allocated for a given image. The output of the descriptor is the set of feature vectors for the initial set of singular points. It should be noted that some descriptors solve two problems simultaneously - find critical points and the construction of these points handles.

Signs (descriptors) are based on the information about the intensity, color and texture of a singular point. But the critical points are presented corners, edges, or contours of objects, therefore, as a rule, calculations are made for a certain neighborhood. Ideally, signs should have a good number of properties:

- Repeatability. The images of the same object or scene taken from different angles and under different lighting conditions, most of the signs should be the detected.
- Locality. Signs should be as local, to reduce the risk of overlap.
- The representativeness. The number of signs should be sufficient to a reasonable number of features detected by even a small image of the object.
- Accuracy. Signs must be accurately detected, in relation to the size and shape of the object.
- Efficiency. For real-time applications is critical to the calculation procedure of signs did not require significant computational costs.

3.2 SIFT

SIFT (Scale Invariant Feature Transform) [19] [20] [21] it identifies and describes the local features of the image. Obtained using a relatively invariant features scale and rotation, resistant to a number of affine transformations, noise.

Algorithm of Scale Invariant Feature Transform is quite effective, but it has a high complexity of processing.This algorithm can be divided into two parts: the definition of "points of interest» (key points, points of interest) and the construction of neighborhoods descriptor data points. There are several ways to determine the points of interest. The algorithm proposed in the framework of SIFT, one of the most famous. It consists in the use of Gaussian pyramid, which is built for the images. Further images are the same size, and calculated the difference between them (DoG, difference-of-Gaussian images), as shown in Figure 10 as well. Moreover, as candidates of points of interest are selected only those pixels that differ from the rest, this is done for example by comparing each pixel with a number of neighboring scale corresponding to several neighbors and more smaller scale, this process is shown in Figure 10, b.

A pixel is selected as the point of interest only if its Brightness is the peak. Next to each of these points of interest calculated local descriptor that characterizes the direction gradients in a neighborhood of pixels (Fig.10).

The pyramid of Gaussians   **Pyramid difference of Gaussians**

a)   b)
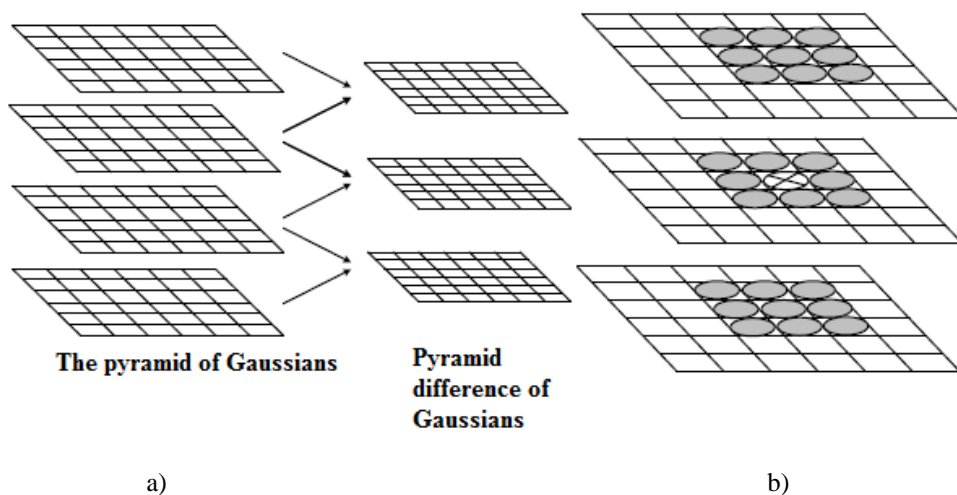
Fig. 10 The process of finding points of interest

The advantage of this method is full of invariance under affine transformations of the four parameters, namely, scale, move and rotate.

SIFT descriptors are not without drawbacks. Not all of the data points and their descriptions will meet the requirements. And it will affect the future solution of the matching images.

The method does not work in the following cases:

- Lighting conditions are different (eg day / night).
- The object has a reflecting surface (usually vehicles, mirrors).
- The object has a distinct 3-D structure.
- The object has no such structure: then across "true" mismatch.
- Angle is too different.

3.3 SURF

An alternative approach is to SURF (Speeded Up Robust Features) [21] [22], which is several times faster than SIFT. In this approach to accelerate the search for points of interest using the integral image. The value at each point in the integral image is calculated as the sum of the values at a given point and the values of all the points that are above and to the left of this. Using the integral images are computed in constant time so-called rectangular filters, which consist of multiple rectangular areas (Fig.11).
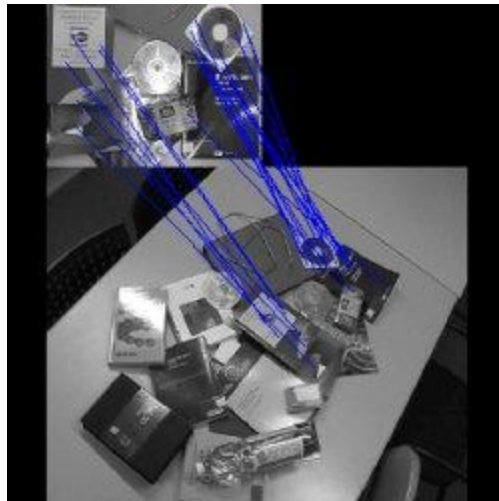
Fig. 11 Example of the Speeded Up Robust Features

SURF solves two problems - find specific points of the image and the creation of descriptors that are invariant to scale and rotation. This means that the description of the key points will be the same, even if the sample size and the change will be rotated.

Besides, the search for points also must have invariance. So that would rotate the object scene had the same set of points as the sample.

Determination of critical points in the image is performed based on Hessian (FAST-Hessian detector [21]). Using Hessian provides invariance under transformation of the "turn", but does not change the scale invariance. Therefore SURF applies filters of all sizes to calculate the Hessian. Suppose that the original image is defined by a matrix of intensities I, the current pixel under consideration is denoted by $X=(x,y),\sigma$ - scale filter. Then the Hessian matrix is of the form (1):

$$H(X,\sigma) = H(X,\sigma) = \begin{pmatrix} L_{xx}(X,\sigma) & L_{xy}(X,\sigma) \\ L_{xy}(X,\sigma) & L_{yy}(X,\sigma) \end{pmatrix} \tag{1}$$

$L_{xx}(X, \sigma)$, $L_{xy}(X, \sigma)$, $L_{yy}(X, \sigma)$ - convolution approximation of the second derivative of a Gaussian kernel with the image. The determinant of the Hessian reaches extremum at the point of maximum change in the gradient of brightness. Therefore SURF iterates filter with a Gaussian kernel of the whole image and find the point at which the maximum value of the determinant of the Hessian. Note that this passage stands out as a dark spot on a white background, and bright spots on a dark background.

Next to each found a singular point is calculated orientation - the direction of the prevailing difference in brightness. The concept of orientation is close to the notion of the direction of the gradient, but to determine the orientation of the singular point Haar filter [23] is applied.

After finding the key points, SURF shapes their descriptions:

- Around the point size of a neighborhood built square 20s, where s - the scale at which the maximum value obtained by the determinant of Hess.
- The resulting square area is divided into blocks, resulting in the region will be divided into 4x4 region.
- For each block calculated more simple features. As a consequence, it is a vector containing four components: 2 - a total gradient quadrant, 2 - the sum of the absolute point gradients.

35

- The handle is formed as a result of bonding descriptions weighted gradient for 16 quadrants around the singular point. Descriptor elements are weighted by coefficients Gaussian kernel. Weight needed for greater noise immunity in remote locations.

- In addition to the descriptor contains the following Hessian. These components are needed to distinguish between light and dark spots. For light dots on dark background footprint is negative, for the dark spots on a light background - positive.

The disadvantage is that SURF used to find objects on the image, he does not work with objects. SURF does not select the object from the background. He treats the image as a whole and looks especially for this image. Thus features may be inside the object, and the background, as well as the boundary points of the object and the background. Therefore, the method does not work well for objects of simple form and texture without pronounced. Inside these objects, the method is not likely to find the singular points. Points will be found either on the border with the background of the object, if any, only in the background. This will lead to the fact that an object can not be recognized in another image on another background. Method weakly sensitive to noise.

Good and stable result is observed during the rotation, changing brightness of the image, or the angle at which the picture was taken. Besides SURF several times faster than SIFT. Good way to cope with the change in brightness of the image.

3.4 MSER.

MSER (Maximally Stable Extremal Regions) [24] [25] [26] [21] techniques and the most invariant to affine transformations and zooming. Both methods normalize 6 parameter affine distortion. Elaborate on MSER.

"Extreme region" - is the name of the method derives from the sort of singular points of the intensity (in the lower and upper levels). To build a pyramid, in which the initial level corresponding to the minimum value of the intensity, is white image, and the last level corresponding to the maximum value of intensity - black. If at some point there is a movement, it is a white image appear black spots corresponding to local minima of intensity. With the increase in the threshold (the threshold value $\tau \in \{0,1 \ldots, 255\}$, pixel intensity is less than the threshold is considered black, otherwise white) spots begin to

grow and merge, eventually forming a single black. This allows you to build a pyramid set of connected components corresponding to the white areas - regions with a maximum intensity. If you invert the binary image in the pyramid, we get a set of regions with a minimum intensity. Chart consists of several steps:

- Sort the set of image pixels in ascending / descending order of intensity. Note that this can be sorted in time proportional to the number of pixels.
- Built a pyramid of connected components. For each pixel of the sorted sets to execute a sequence of actions:
  - update the list of points that make up the components;
  - update areas following components, whereby the pixels of the previous level will be a subset of pixels of the next level.
- Run to find all the components of local minima (find pixels that are present in this component, but not part of the previous ones). A set of local minima correspond to the level of extreme regions of the image.

Lack MSER that he did not fully invariant to scale. In this method the normalization scale occurs depending on the size of the extreme regions. However, scaling is not the same, it includes region fuzzy borders

the following sub-sample. Blur brings together the region and changes their shape and size. In other words, the method is not restricted commutes between the optical blur and affine transformation.

The function of the sparsity, is another negative side MSER. This method uses only a very contrasting levels of the set, however, experts suggest that MSER resistant to change in the slope τ_{max} from 5 to 10, and higher performance than SIFT. But this condition is satisfied when there is no significant change of scale between the images, and if the image contains enough contrast objects.

But we must not lose sight of the benefits and MSER. As specific areas determined exclusively function of the intensity of the image, it results in many key performance areas, which makes them useful. Detected area have the following properties:

- Invariance to affine transformations.
- Saving covariance when converting image areas.
- Stability: addressing areas only within the selected range of extreme.

o   Multi-scale, the detection areas as small and large structures.

MSER - a method that normalizes all 6 parameter affine transformations.

3.5 Harris-Affine and Hessian-Affine.

Harris-Affine [27] [26] [20] [21] [28] and Hessian-Affine [27] normalize the parameters of affine transformations. Harris, as special areas using angles and identify the key points in the scale space, using the approach proposed by Lindenberg (Fig.12).
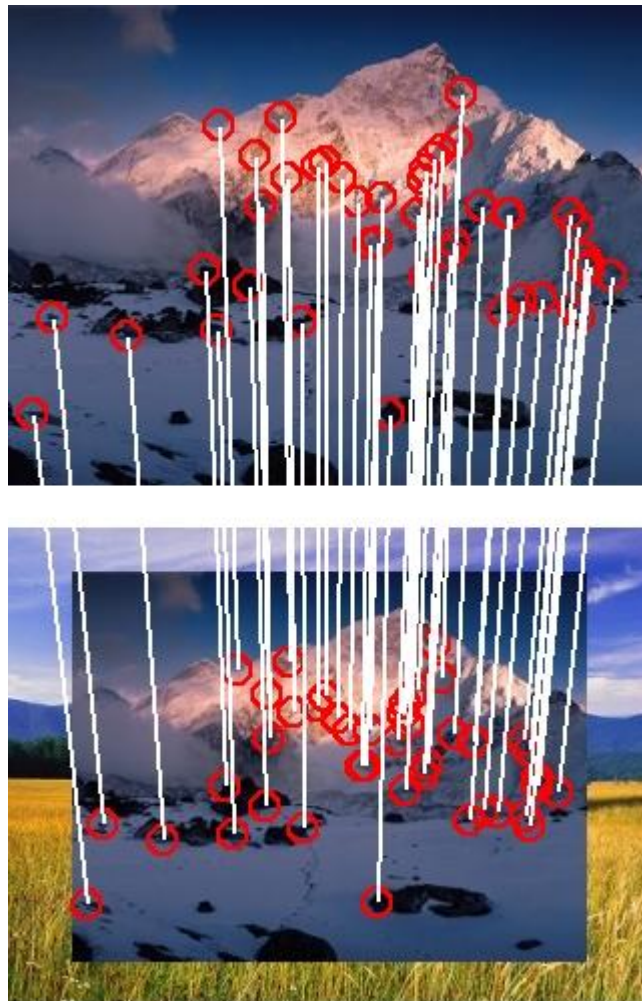

Fig. 12 Example of the Harris-Affine and Hessian-Affine

Affine normalization is carried out by repeating the procedure to estimate the parameters of the elliptic region and normalize them. With each repetition of the elliptic region occurs parameter estimation: minimized the difference between the intrinsic angular momentum of second order matrices selected area; elliptical area is normalized to a circular; It carried out assessment of the key points of its scale in the scale of the space.

This procedure is repeated homogeneous areas that are covariant with respect to affine transformations. Own points of second order matrices are used to measure affinity form spot area.

Hessian-Affine similar to the Harris-Affine, but uses as a special area of blobs instead of angles. The determinant of the Hessian local maxima is used as reference points. The remainder of the method is the same as the Harris-Affine.

As noted in both methods, the first step - a lot of the enormity of Harris and Hessian detectors. Features on the results of repeated procedures should lead to an affine invariant.

Experts indicate that the Hessian-Affine and Harris-Affine resistant to change in the slope $\tau_{max} \approx 2.5$. This disappointing result can be explained by the repeated failure of the procedure capture large tilt.

Hessian-Affine detector works well with the texture areas with images that have a large number of angular units. Very good used in some structured scenes, such as buildings.

Harris Detector partially invariant to changes in illumination, Pan and Zoom. It remains very stable at blur]. Because the detector is independent of image segmentation, repeatability and corresponding estimates remain unchanged.

The disadvantage of these methods is that they are poorly resistant to change the viewing angle and zoom.

3.6 Affine-SIFT (ASIFT).

ASIFT - the idea of unification and normalization of the main parts of the method of SIFT. SIFT detector normalizes rotation, moving and simulates all deleted from the search and query image(Fig.13).
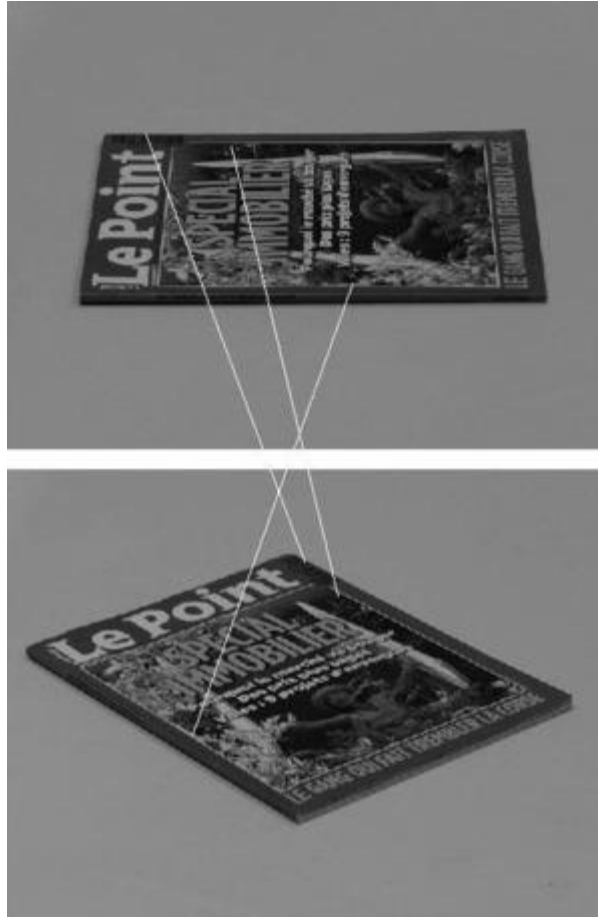
Fig. 13 Example of the ASIFT

ASIFT models with sufficient accuracy all the distortions caused by the change in the position of optical cameras. Then, the method SIFT.Drugimi words ASIFT method simulates three parameters: scale, longitude angle camera and the latitude angle (which is equivalent to the slope) and normalizes other parameters: translation and rotation.

A key feature is that the slope of the distortion is irreversible because noncommutativity with blurred outlines, it may be offset by scaling up a digital simulation of slope.

ASIFT main advantage is that it simulates all distortions caused by the optical camera (scale, angle, tilt angle). The method is invariant to translation, rotation and scale up to full affine invariant.

The drawback is that the affine transformation are stored other disadvantages of the method SIFT: refusal to operate at different light at a reflecting surface, etc.

All of the above local descriptors share the same paradigm is to find points of interest and building for each descriptor that describes the neighborhood.

In this field, the development of methods considered, progress has been made. In a comparative analysis revealed that none of the methods was not without flaws. Scale and rotation invariant method ASIFT and SIFT. To change the camera position and stable MSER ASIFT. When fuzzy boundaries and work well ASIFT SIFT, less stable methods Hessian-Affine and Harris-Affine. In fact, the only method that meets all requirements and achieves maximum affine invariant is ASIFT, but it does have its drawbacks.

3.7 GLOH

Descriptor GLOH (Gradient location-orientation histogram) [29] is a modification SIFT-descriptor, which is built in order to improve reliability. Upon SIFT descriptor is calculated, but the polar grid is used to partition the neighborhood bins (Fig. 14): three radial block with a radius of 6, 11 and 15 pixels, and 8 sectors. The result is a vector containing the components 272, which is projected in a space of dimension 128 by using principal component analysis (PCA) (Fig.14).
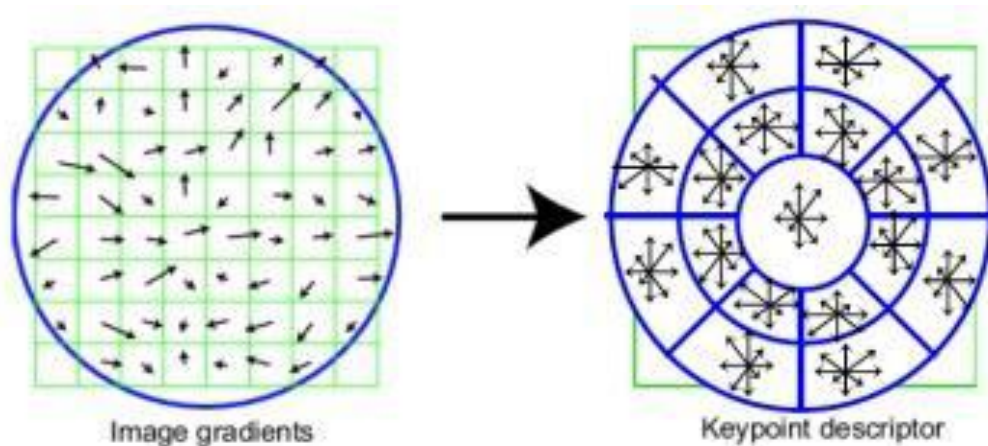
Fig. 14 Polar Grid division into bins

3.8 Conclusion

In this chapter we considered different types of the descriptors and detectors which are used in computer vision. At this stage we finished with theoretical part, and we can start to practical, namely calibrations of the camera.

41

4 Camera Calibration

4.1   Reading the images

First of all we need to calibrate camera. Calibration of the camera — is a problem of obtaining internal and external parameters of the camera of the available photos or videos, finished shooting by it.

Distortion (from the Latin. distorsio, distortio - bending) - an aberration of optical systems, in which a linear increase varies across the field of view. In this case it violated the geometric similarity between the object and its image. Distortion is unacceptable in optics designed for aerial photography and photogrammetric production of photomasks. The optical system is free from distortion, called orthoscopic [30] (Fig. 15).
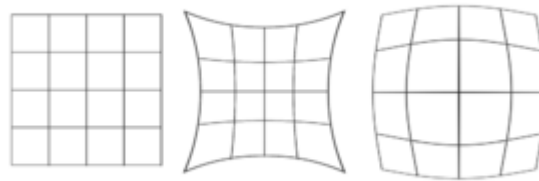
Fig. 15 Distortion of the square (left) in optical systems with a negative (center) and positive (right) distortion

It can be corrected by selection of lenses and other optical elements of the system during its development. The least pronounced in lenses symmetrical design. In digital photography and cinematography distortion can be corrected by image processing in the computer.

The aberration of the optical system - an error in the optical system of the image caused by deflection of the beam from the direction in which he would have to go to an ideal optical system.

Calibration of the camera is often used at the initial stage of the solution of many problems of computer sight and in particular augmented reality. Besides, calibration of the camera helps to correct a distortion in photos and video. The camera isn't ideal the optical device, especially for technical sight therefore the removed image is required to be corrected at first that is the elementary optical distortions. We will do this with camera calibration toolbox for matlab [31].

Camera model parameters

As a rule, for representation 2D - point coordinates on the plane the look vector column $[u\ v\ 1]^T$, and for a situation 3D task - a point in world coordinates $—[x_w y_w z_w 1]^T$. It should be noted that these expressions are written down in the expanded notation of uniform coordinates which is the most widespread in a robotics and problems of transformation of solid bodies. In particular, in model of a camera obscura the matrix of the camera is used for projection of points of three-dimensional space on the image plane (2):

$$z_c \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = A(R \quad T) \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} \tag{2}$$

where z_c - any large-scale coefficient

Parameters of internal calibration

$$A = \begin{pmatrix} \alpha_x & \gamma & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3}$$

The matrix of internal calibration of A (3) contains 5 significant parameters. These parameters correspond to focal length, a tilt angle of pixels and a basic point (the image plane point of intersection with an optical axis coinciding with the center of the photo. In real cameras it is, as a rule, a little displaced because of optical distortions). In particular, α_x and α_y correspond to the focal length measured in width and height of pixel u_0 and v_0 — coordinates of a basic point, γ=ay*tanϕ, where ϕ - pixel tilt angle [32]. Nonlinear parameters of internal calibration, such as distortion coefficients, are also important, though can't be included in the linear model described by a matrix of internal calibration. The majority of modern algorithms of calibration of the camera defines them together with parameters of linear part of model.

Parameters of internal calibration belong only to the camera, but not to a scene therefore they change only in that case when the corresponding settings of the camera change.

Parameters of external calibration

R, T (where R — a vector 1×3 or a matrix 3×3 turns, T — a vector 3×1 transfers) — the parameters of external calibration defining the transformation of coordinates transferring coordinates of points of a scene from world system of coordinates to the system of coordinates connected with a chamber [32]. Or that is equivalent to the previous definition, parameters of external calibration set position of the camera in world system of coordinates.

Parameters of external calibration are connected directly with the photographed scene therefore (unlike parameters of internal calibration) to each photo there corresponds the set of these parameters.

First, we need to read images taken with the camera in matlab camera calibration toolbox. Now we have 30 images of a planar checkerboard(Fig.16).
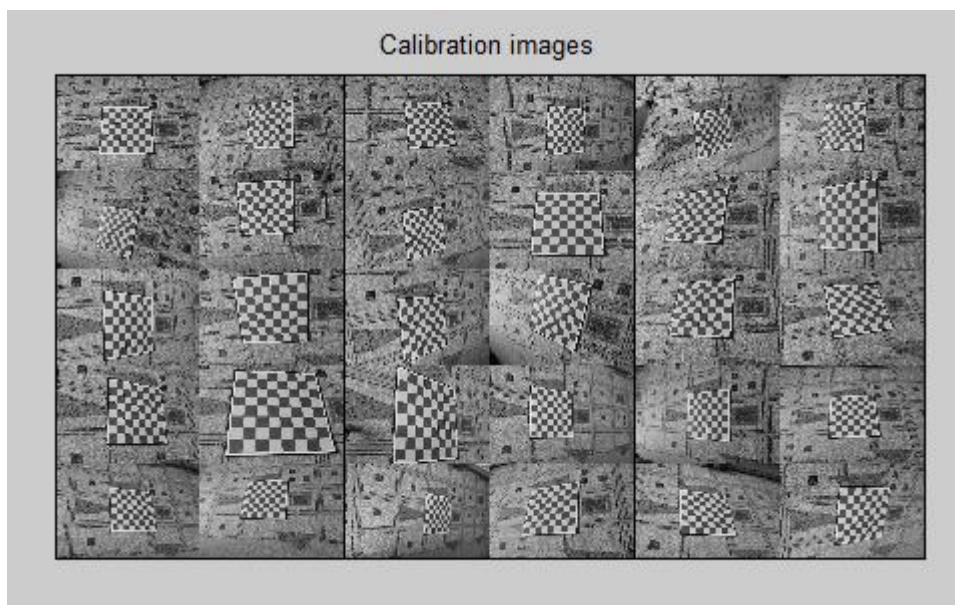
Fig. 16 The complete set of images

4.2   Extract the grid corners:

Ordering rule for clicking: The first clicked point is selected to be associated to the origin point of the reference frame attached to the grid. The other three points of the

rectangular grid can be clicked in any order. This first-click rule is especially important if you need to calibrate externally multiple cameras (i.e. compute the relative positions of several cameras in space). When dealing with multiple cameras, the same grid pattern reference frame needs to be consistently selected for the different camera images (i.e. grid points need to correspond across the different camera views)(Fig.17).
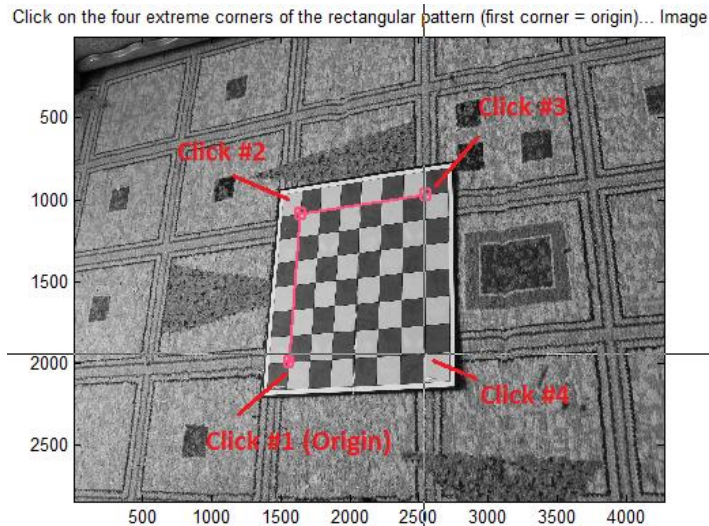

Fig. 17 Ordering rule for clicking

The boundary of the calibration grid is then shown (Fig. 18):
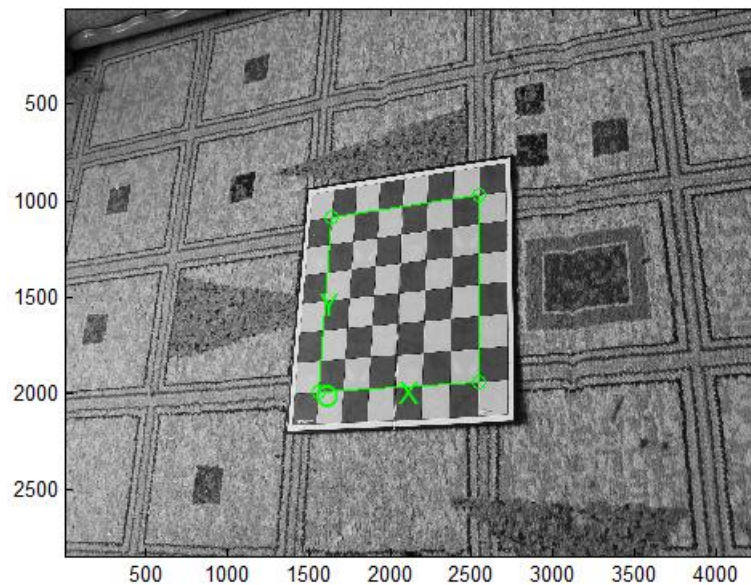
Fig. 18 The boundary of the calibration grid

The image corners are then automatically extracted, and displayed (the blue squares around the corner points show the limits of the corner finder window) (Fig. 19):
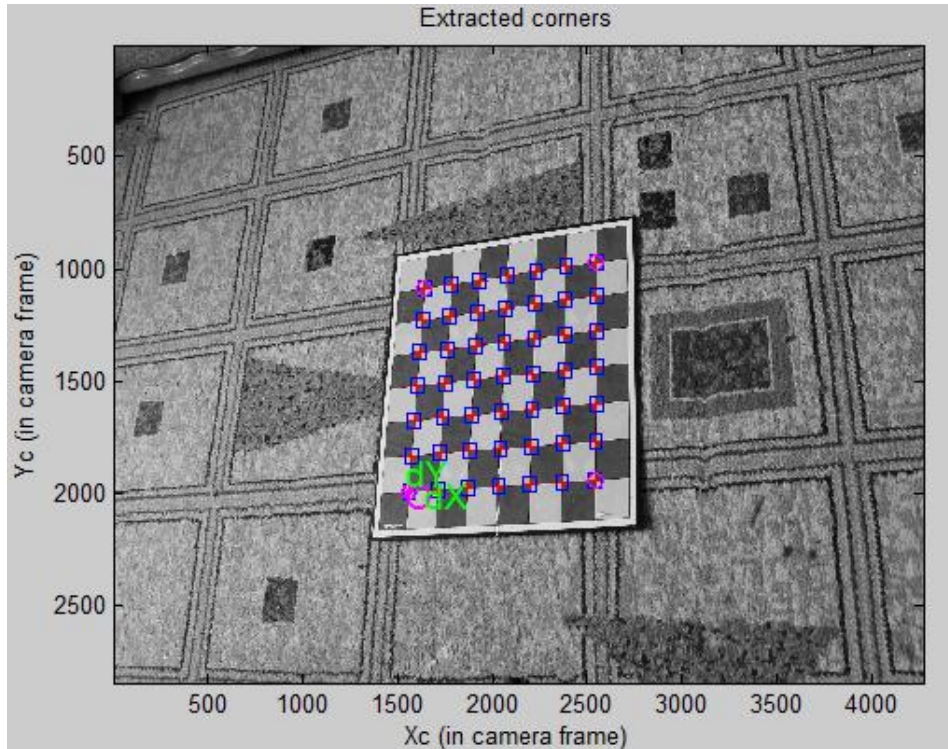
Fig. 19 The image corners

4.3   Main Calibration step

After all calibration we have the reprojection error wich shown in the form of color-coded crosses(Fig.20).

The **reprojection error** is a geometric error corresponding to the image distance between a projected point and a measured one. It is used to quantify how closely an estimate of a 3D point $\hat{X}$ recreates the point's true projection x. More precisely, let P be the projection matrix of a camera and $\hat{x}$ be the image projection of $\hat{X}$, i.e. $\hat{x} = P\hat{X}$. The reprojection error of $\hat{X}$ is given by $d(x, \hat{x})$, where $d(x, \hat{x})$, denotes the Euclidean distance between the image points represented by vectors x and $\hat{x}$.

Minimizing the reprojection error can be used for estimating the error from point correspondences between two images. Suppose we are given 2D to 2D point imperfect correspondences $\{x_i \leftrightarrow \acute{x}_i\}$. We wish to find a homography $\hat{H}$ and pairs of perfectly matched points $\hat{x}_i$ and $\hat{\hat{x}}_i$, i.e. points that satisfy $\hat{\hat{x}}_i = \hat{H}\hat{x}_i$ that minimize the reprojection error function given by (4)

46

$$\sum_i d(x_i, \hat{x}_i)^2 + d(\acute{x}_i, \hat{\acute{x}}_i)^2 \qquad (4)$$

So the correspondences can be interpreted as imperfect images of a world point and the reprojection error quantifies their deviation from the true image projections $\hat{x}_i, \hat{\acute{x}}_i$
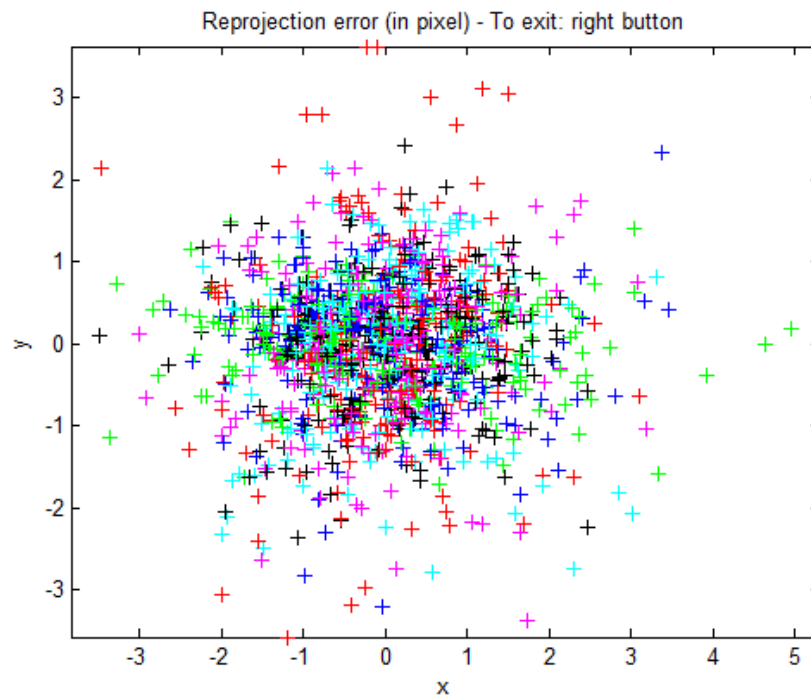

Fig. 20 Reprojection error

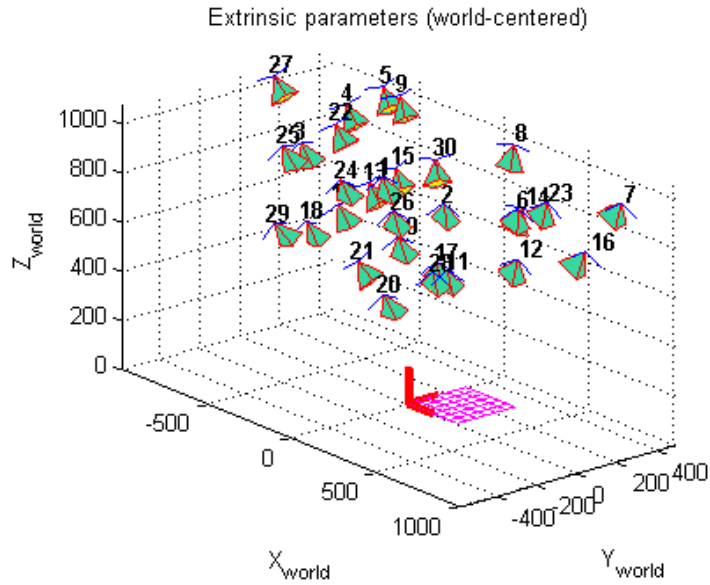On this new figure(Fig.21), every camera position and orientation is represented by a green pyramid.

Fig. 21 Camera positions

All parameters were saved in matlab file **Calib_Results.mat**.

Calibration results after optimization (with uncertainties):

Focal Length: fc $=$ [4178.76736 4185.21459] $\pm$ [15.19170 15.59226]

Principal point: cc $=$ [2191.16899 1412.23350] $\pm$ [19.81314 21.05832]

Skew: alpha_c $=$ [0.00000] $\pm$ [0.00000] $=>$ angle of pixel axes $=$ 90.00000 $\pm$ 0.00000 degrees

Distortion: kc $=$ [-0.13172 0.21292 -0.00057 0.00261 0.00000] $\pm$ [0.02043 0.19538 0.00133 0.00132 0.00000]

Pixel error: err $=$ [1.06892 0.84622]

Note: The numerical errors are approximately three times the standard deviations (for reference).

4.4   Conclusion

In this chapter we calibrated the camera using camera calibration toolbox for matlab After calibration of the camera we can start the next step, finding the relationship between two images.

48

5 Experement

5.1   Introduction

After calibration we can start an experiment. For a start it is necessary to download and specify a way to libraries:

- Robotic toolbox
- Machine vision toolbox
- Math toolbox

After that we can associate camera coordinates and coordinates of the frame. First step will be reading photos and matching them, to find the relationship between them. Next step will be finding matrix of homography. And at last find associating between coordinates.

5.2   Matching photos and finding the matrix of homography.

As a descriptor has been selected Harris corner detector. We need to find this descriptors because descriptor — the identifier of a key point allocating it from other mass of special points. In turn, descriptors have to provide invariancy of finding of compliance between special points concerning transformations of images. In principle, as a descriptor any object (if only he coped with the functions) can act, but usually a descriptor is certain information on the vicinity of a key point. Such choice is made owing to several reasons: effects of distortions, some changes have smaller impact on small areas (change of provision of object on the picture, change of a scene, overlapping of one object by another, turn) can not affect a descriptor at all. First, we read images and find the best corner points on image.

Iread read image from file

im = iread() presents a file selection GUI from which the user can select an image file which is returned as a matrix. On subsequent calls the initial folder is as set on the last call.

im = iread ([], options) as above but allows options to be specified.

im = iread (path, options) as above but the GUI is set to the folder specified by path. If the path is not absolute it is searched for on the matlab search path.

im = iread (file, options) reads the specified image file and returns a matrix. If the path is not absolute it is searched for on matlab search path.

The image can be greyscale or color in any of a wide range of formats supported by the matlab imread function.

Option 'double' - return an image with double precision floating point pixels in the range 0 to 1.

Wildcards are allowed in file names. If multiple files match a 3D or 4D image is returned where the last dimension is the number of images in the sequence.

Notes:

- A greyscale image is returned as an HxW matrix
- A color image is returned as an HxWx3 matrix
- A greyscale image sequence is returned as an HxWxN matrix where N is the sequence length
- A color image sequence is returned as an HxWx3xN matrix where N is the sequence length

Icorner corner detector

f = icorner(im, options) is a vector of PointFeature objects describing corner features detected in the image im. This is a non-scale space detector and by default the Harris method is used but Shi-Tomasi and Noble are also supported.

If IM is an image sequence a cell array of PointFeature vectors for the correspnding frames of IM.

The PointFeature object has many properties including:
u   horizontal coordinate
v   vertical coordinate
strength   corner strength
descriptor   corner descriptor (vector)

Options:

'nfeat',N   return the N strongest corners (default Inf)

'color'   specify that im is a color image not a sequence

'patch',P   use a PxP patch of surrounding pixel values as the feature vector. The vector has zero mean and unit norm.

Notes:

- Corners are processed in order from strongest to weakest.
- The function stops when:
    - the corner strength drops below cmin, or
    - the corner strength drops below cMinThresh x strongest corner, or
    - the list of corners is exhausted
- Features are returned in descending strength order
- If im has more than 2 dimensions it is either a color image or a sequence
- If im is NxMxP it is taken as an image sequence and F is a cell array whose elements are feature vectors for the corresponding image in the sequence.
- If im is NxMx3 it is taken as a sequence unless the option 'color' is given
- If im is NxMx3xP it is taken as a sequence of color images and F is a cell array whose elements are feature vectors for the corresponding color image in the sequence.
- The default descriptor is a vector [Ix* Iy* Ixy*] which are the unique elements of the structure tensor, where * denotes squared and smoothed.
- The descriptor is a vector of float types to save space

After that we will find a vector of point feature objects describing corner features detected in the image this corner points and display the result.(Fig.22)

Idisp Interactive image display tool

Idisp(im, options) displays an image and allows interactive investigation of pixel values, linear profiles, histograms and zooming. The image is displayed in a figure with a toolbar across the top. If im is a cell array of images, they are first concatenated (horizontally).

Notes:

- Is a wrapper around the matlab builtin function image. See the matlab help on "Display Bit-Mapped Images" for details of color mapping.
- Color images are displayed in matlab true color mode: pixel triples map to display RGB values. (0,0,0) is black, (1,1,1) is white.
- Greyscale images are displayed in indexed mode: the image pixel value is mapped through the color map to determine the display pixel value.
- For grey scale images the minimum and maximum image values are mapped to the first and last element of the color map, which by default ('greyscale') is the range black to white. To set your own scaling between displayed grey level and pixel value use the 'cscale' option.
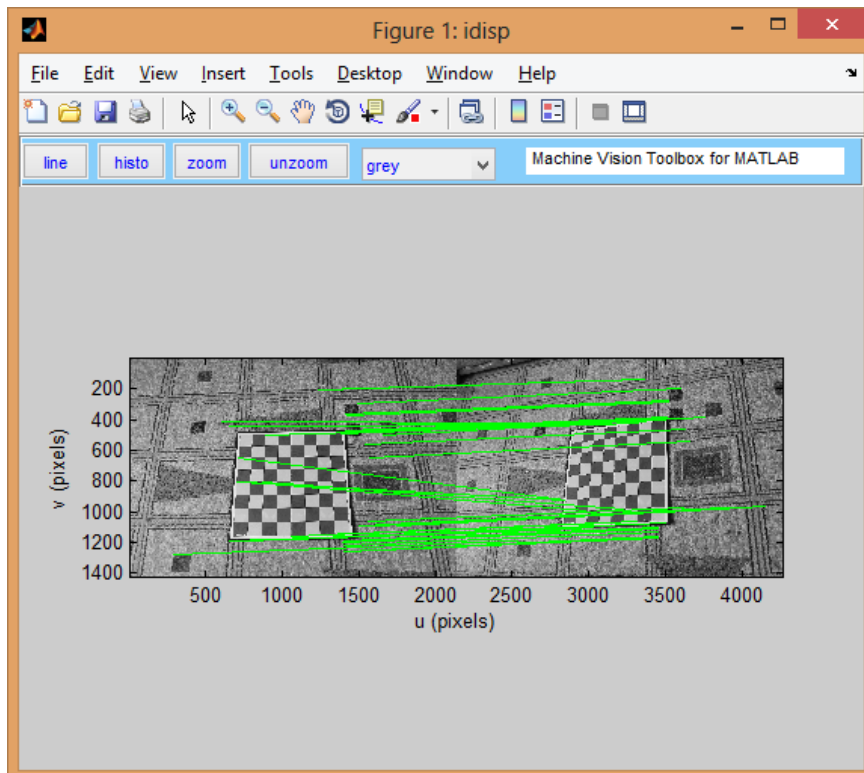
Fig. 22 The result of matching

As you see on the previous figure some points didn't match. It is caused by some error. This error has been received due to the erroneous finding descriptors. Then we can construct a histogram of error.(Fig.23.)

Ihist Image histogram

Ihist(im, options) displays the image histogram. For an image with multiple planes the histogram of each plane is given in a separate subplot.

H = ihist(im, options) is the image histogram as a column vector. For an image with multiple planes H is a matrix with one column per image plane.

Option:

'normcdf'   compute a normalized cumulative histogram, whose maximum value is one.

Notes:

- For a uint8 image the MEX function FHIST is used (if available)
  - The histogram always contains 256 bins
  - The bins spans the greylevel range 0-255.
- For a floating point image the histogram spans the greylevel range 0-1.
- For floating point images all NaN and Inf values are first removed.
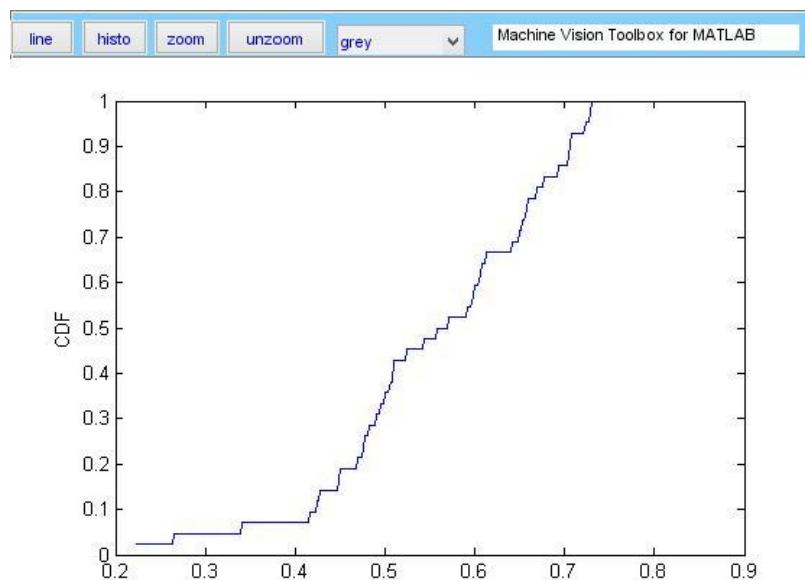
Fig. 23 Histogram of error

Then we can to translate the matrix and display approximate position of the camera. Next figure will show position of the camera associating with the first photoshot and position of the camera associating with the second photoshot (Fig.24.)

53

Transl Create translational transform

T = transl(X, Y, Z) is a homogeneous transform representing a pure translation.

T = transl(P) is a homogeneous transform representing a translation or point P=[X,Y,Z]. If P (Mx3) it represents a sequence and T (4x4xM) is a sequence of homogenous transforms such that T(:,:,i) corresponds to the i'th row of P.

P = transl(T) is the translational part of a homogenous transform as a 3-element column vector. If T (4x4xM) is a homgeneous transform sequence the rows of P (Mx3) are the translational component of the corresponding transform in the sequence.

Notes:

- Somewhat unusually this function performs a function and its inverse.

CentralCamera Perspective camera class

A concrete class for a central-projection perspective camera, a subclass of Camera.

This camera model assumes central projection, that is, the focal point is at z=0 and the image plane is at z=f. The image is not inverted.

Notes:

- Camera is a reference object.
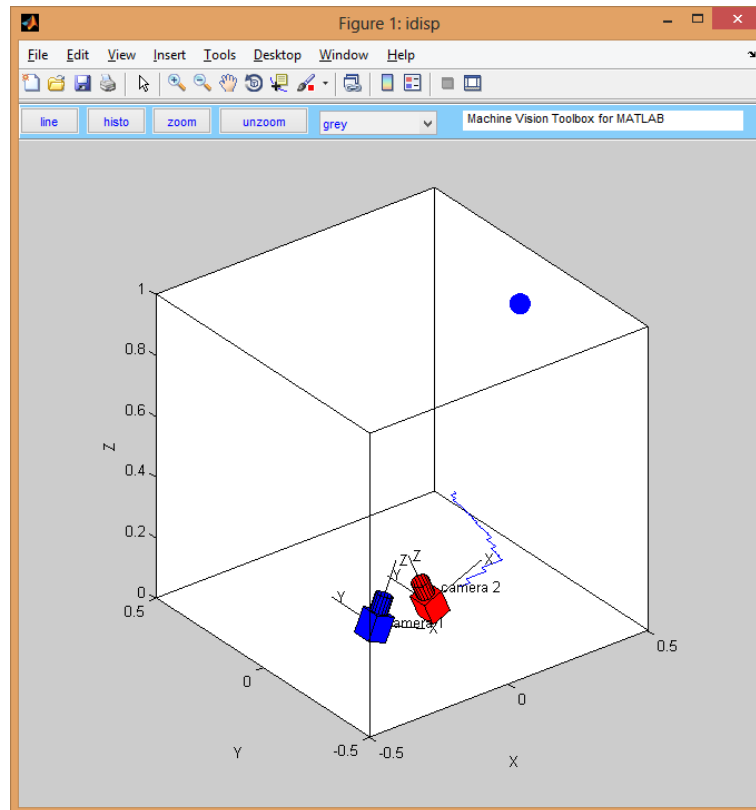- Camera objects can be used in vectors and arrays

Fig. 24 Approximate position of the camera

Plots the epipolar lines due to the fundamental matrix F and the image points p (Fig.25.). Epipolar lines are needed in order to approximate matching descriptors. Nearest to the desired result, matches are acceptable. Those coincidences that are further away from the desired result - an error.

H2e Homogeneous to Euclidean

E = h2e(H) is the Euclidean version (K-1xN) of the homogeneous points H (KxN) where each column represents one point in P^K.
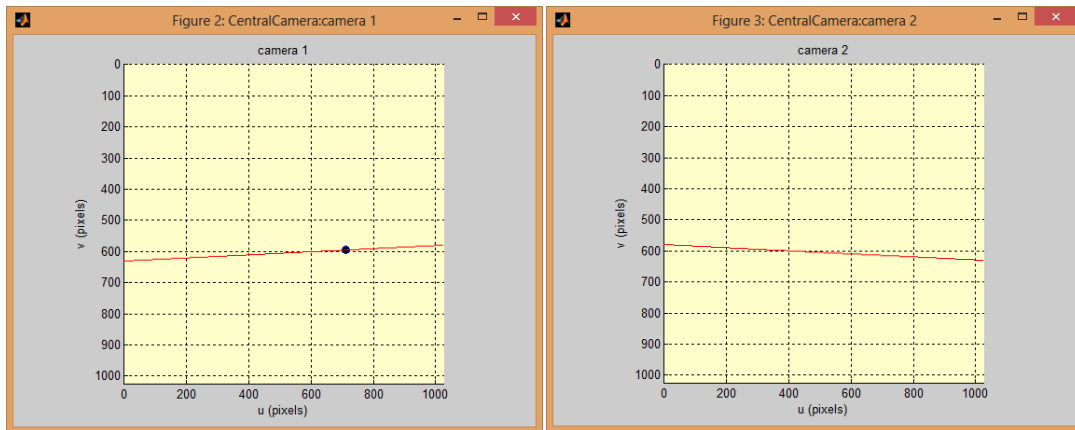
Fig. 25 Plots the epipolar lines

And through all points. On the next figure you can see the scatter of the epipolar lines of the descriptors. (Fig.26).
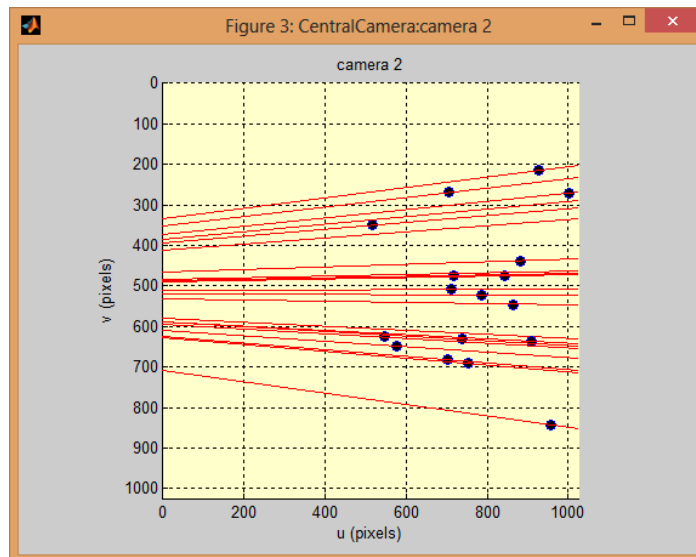

Fig. 26 Plots the epipolar lines through all points

Next step to find the smallest distance of point from epipolar line.(Fig.27)

Fmatrix Estimate fundamental matrix

F = fmatrix(P1, P2, options) is the fundamental matrix (3x3) that relates two sets of corresponding points P1 (2xN) and P2 (2xN) from two different camera views.

Notes:

- The points must be corresponding, no outlier rejection is performed.
- Contains a RANSAC driver, which means it can be passed to ransac().

- F is a rank 2 matrix, that is, it is singular.

Epidist Distance of point from epipolar line

D = epidist(F, P1, P2) is the distance of the points P2 (2xM) from the epipolar lines due to points P1 (2xN) where F (3x3) is a fundamental matrix relating the views containing image points P1 and P2.

D (NxM) is the distance matrix where element D(i,j) is the distance from the point P2(j) to the epipolar line due to point P1(i).

RANSAC Random sample and consensus

M = ransac(FUNC, X, T, options) is the RANSAC algorithm that robustly fits data X to the model represented by the function FUNC. RANSAC classifies Points that support the model as inliers and those that do not as outliers.

X typically contains corresponding point data, one column per point pair. RANSAC determines the subset of points (inliers) that best fit the model described by the function FUNC and the parameter M. T is a threshold on how well a point fits the estimated, if the fit residual is aboe the the threshold the point is considered an outlier [33] [34].

[M,IN] = ransac(FUNC, X, T, options) as above but returns the vector IN of column indices of X that describe the inlier point set.

[M,IN,RESID] = ransac(FUNC, X, T, options) as above but returns the final residual of applying FUNC to the inlier set.

Notes:

- For some algorithms (eg. fundamental matrix) it is necessary to condition the data to improve the accuracy of model estimation. For efficiency the data is conditioned once, and the data transform parameters are kept in the .misc element. The inverse conditioning operation is applied to the model to transform the estimate based on conditioned data to a model applicable to the original data.
- The functions FMATRIX and HOMOG are written so as to be callable from RANSAC, that is, they detect a structure argument.
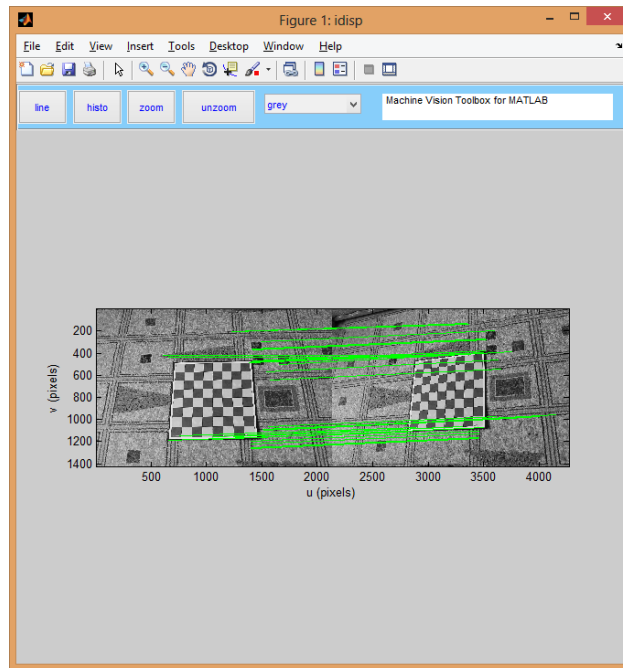
Fig. 27 The smallest distance of point

Also we can display our error detection point matches to look at points which don't satisfy our conditions. (Fig.28).
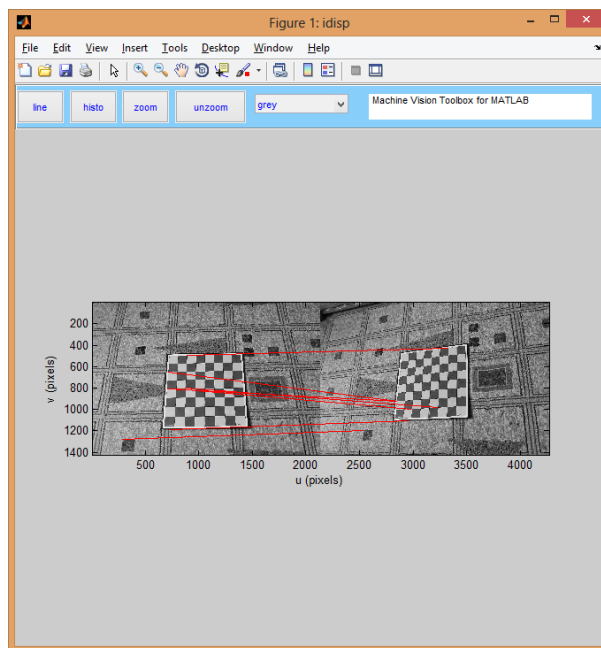

Fig. 28 Error detection point matches

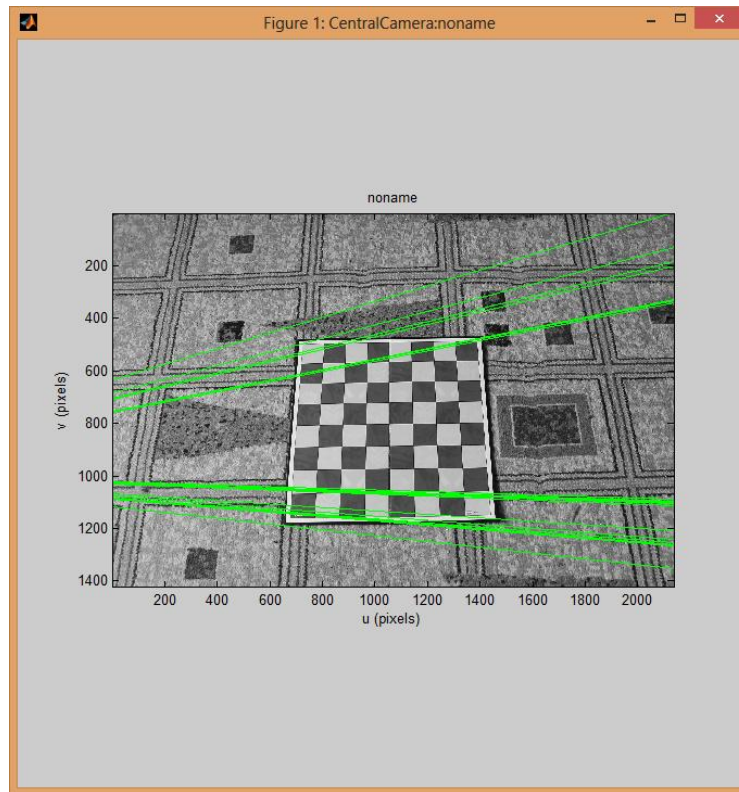Then we make homogeneous to Euclidean(Fig.29)

Fig. 29 Homogeneous to Euclidean

Create grid of points(Fig.30)

function G = mkgrid(M, N, doPlot)

Create an M by N undirected grid as an adjacency matrix in column order i.e. mkgrid(2, 2) creates a grid that looks like this:
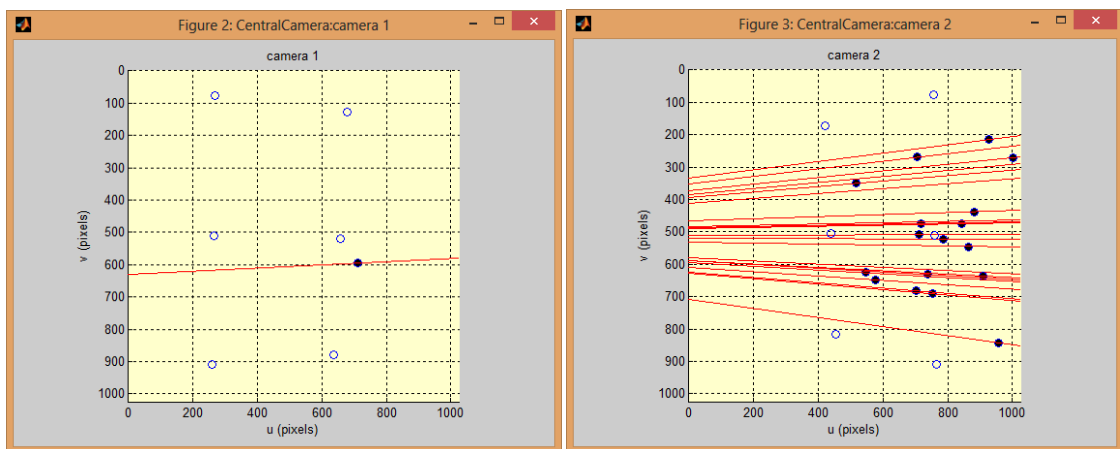
$$\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$$

Fig. 30 Grid points

After that we build matrix of homography and display cam2 (Fig.31).

Homography Estimate homography

H = homography(P1, P2) is the homography (3x3) that relates two sets of corresponding points P1 (2xN) and P2 (2xN) from two different camera views of a planar object.

Notes:

- The points must be corresponding, no outlier rejection is performed.
- The points must be projections of points lying on a world plane
- Contains a RANSAC driver, which means it can be passed to ransac().

Homtrans Apply a homogeneous transformation

P2 = homtrans(T, P) applies homogeneous transformation T to the points stored columnwise in P.

- If T is in SE(2) (3x3) and
    - P is 2xN (2D points) they are considered Euclidean (R^2)
    - P is 3xN (2D points) they are considered projective (P^2)
- If T is in SE(3) (4x4) and
    - P is 3xN (3D points) they are considered Euclidean (R^3)
    - P is 4xN (3D points) they are considered projective (P^3)

TP = homtrans(T, T1) applies homogeneous transformation T to the homogeneous transformation T1, that is TP=T*T1. If T1 is a 3-dimensional transformation then T is applied to each plane as defined by the first two dimensions, ie. if T = NxN and T=NxNxP then the result is NxNxP.
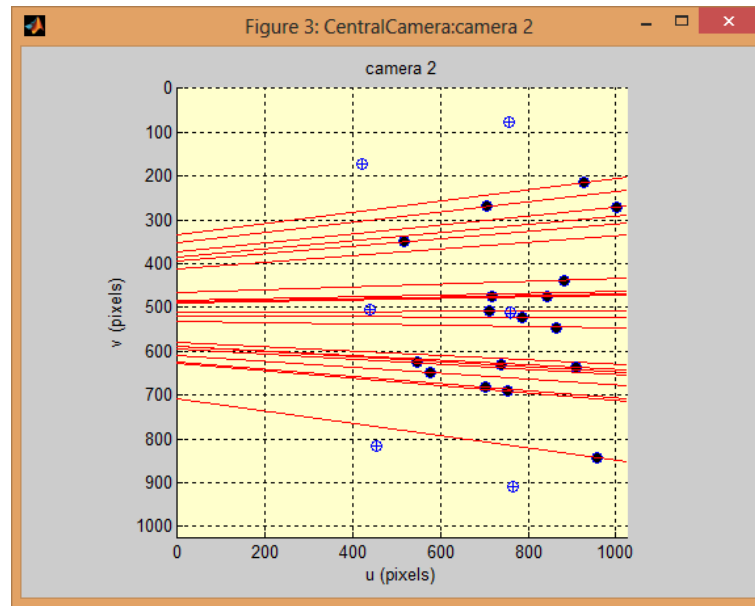
Fig. 31 Matrix of homography and display cam2

Next we display the position of the camera relative frame(Fig.32).

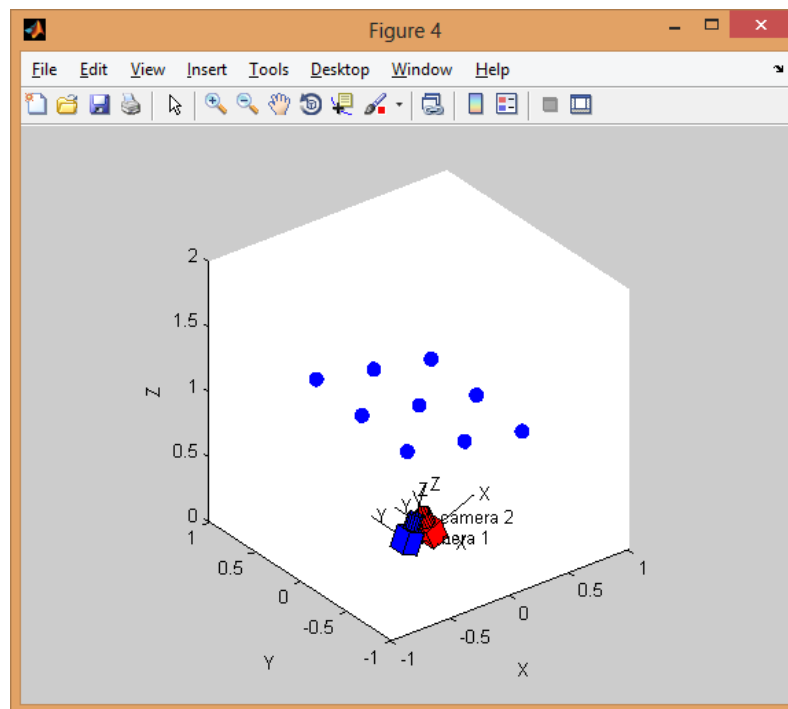

Fig. 32 The position of the camera relative frame

5.3 Associating Coordinates

The last a step of this work is to connect coordinates of the camera and coordinate of the image.

We have matrix of homography and parameters of camera.

The list of internal parameters:

- **Focal length:** The focal length in pixels is stored in the 2x1 vector **fc**.

- **Principal point:** The principal point coordinates are stored in the 2x1 vector **cc**.

- **Skew coefficient:** The skew coefficient defining the angle between the x and y pixel axes is stored in the scalar **alpha_c**.

- **Distortions:** The image distortion coefficients (radial and tangential distortions) are stored in the 5x1 vector **kc**.

Definition of the intrinsic parameters:

Let **P** be a point in space of coordinate vector $\mathbf{XX_c} = [\mathbf{X_c};\mathbf{Y_c};\mathbf{Z_c}]$ in the camera reference frame.

Let us project now that point on the image plane according to the intrinsic parameters (**fc,cc,alpha_c,kc**).

Let $\mathbf{x_n}$ be the normalized (pinhole) image projection (5):

$$x_n = \begin{pmatrix} Xc/Zc \\ Yc/Zc \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \tag{5}$$

Let $r^2 = x^2 + y^2$

After including lens distortion, the new normalized point coordinate $\mathbf{x_d}$ is defined as follows (6):

$$x_d = \begin{pmatrix} x_d(1) \\ x_d(2) \end{pmatrix} = (1 + kc(1)r^2 + kc(2)r^4 + kc(5)r^6)x_n + dx \tag{6}$$

where **dx** is the tangential distortion vector (7):

$$dx = \begin{pmatrix} 2kc(3)x\,y + kc(4)(r^2 + 2x^2) \\ kc(3)(r^2 + 2y^2) + 2kc(4)x\,y \end{pmatrix} \tag{7}$$

Therefore, the 5-vector **kc** contains both radial and tangential distortion coefficients (observe that the coefficient of 6th order radial distortion term is the fifth entry of the vector **kc**).

It is worth noticing that this distortion model was first introduced by Brown in 1966 and called "Plumb Bob" model (radial polynomial + "thin prism"). The tangential distortion is due to "decentering", or imperfect centering of the lens components and other manufacturing defects in a compound lens. For more details [35].

Once distortion is applied, the final pixel coordinates **x_pixel** = **[x_p;y_p]** of the projection of **P** on the image plane is (8):

$$\begin{cases} x_p = fc(1)\big(x_d(1) + alpha_c * x_d(2)\big) + cc(1) \\ \qquad y_p = fc(2)x_d(2) + cc(2) \end{cases} \tag{8}$$

Therefore, the pixel coordinate vector **x_pixel** and the normalized (distorted) coordinate vector **x_d** are related to each other through the linear equation (9):

$$\begin{pmatrix} x_p \\ y_p \\ 1 \end{pmatrix} = KK \begin{pmatrix} x_d(1) \\ x_d(2) \\ 1 \end{pmatrix} \tag{9}$$

where **KK** is known as the camera matrix, and defined as follows (10):

$$KK = \begin{pmatrix} fc(1) & alpha_c * fc(1) & cc(1) \\ 0 & fc(2) & cc(2) \\ 0 & 0 & 1 \end{pmatrix} \tag{10}$$

In matlab, this matrix is stored in the variable **KK** after calibration. Observe that **fc(1)** and **fc(2)** are the focal distance (a unique value in mm) expressed in units of horizontal and vertical pixels. Both components of the vector **fc** are usually very similar. The ratio **fc(2)/fc(1)**, often called "aspect ratio", is different from 1 if the pixel in the CCD

array are not square. Therefore, the camera model naturally handles non-square pixels. In addition, the coefficient **alpha_c** encodes the angle between the x and y sensor axes. Consequently, pixels are even allowed to be non-rectangular. Some authors refer to that type of model as "affine distortion" model.

In addition to computing estimates for the intrinsic parameters **fc**, **cc**, **kc** and **alpha_c**, the toolbox also returns estimates of the uncertainties on those parameters. The matlab variables containing those uncertainties are **fc_error**, **cc_error**, **kc_error**, **alpha_c_error**. For information, those vectors are approximately three times the standard deviations of the errors of estimation.

Extrinsic parameters:

- **Rotations:** A set of **n_ima** 3x3 rotation matrices **Rc_1**, **Rc_2**,.., **Rc_20** (assuming **n_ima=20**).

- **Translations:** A set of **n_ima** 3x1 vectors **Tc_1**, **Tc_2**,.., **Tc_20** (assuming **n_ima=20**).

Definition of the extrinsic parameters:

Consider the calibration grid **#i** (attached to the **i**th calibration image), and concentrate on the camera reference frame attahed to that grid.

Without loss of generality, take **i = 1**. The following figure shows the reference frame **(O,X,Y,Z)** attached to that calibration grid (Fig.33).
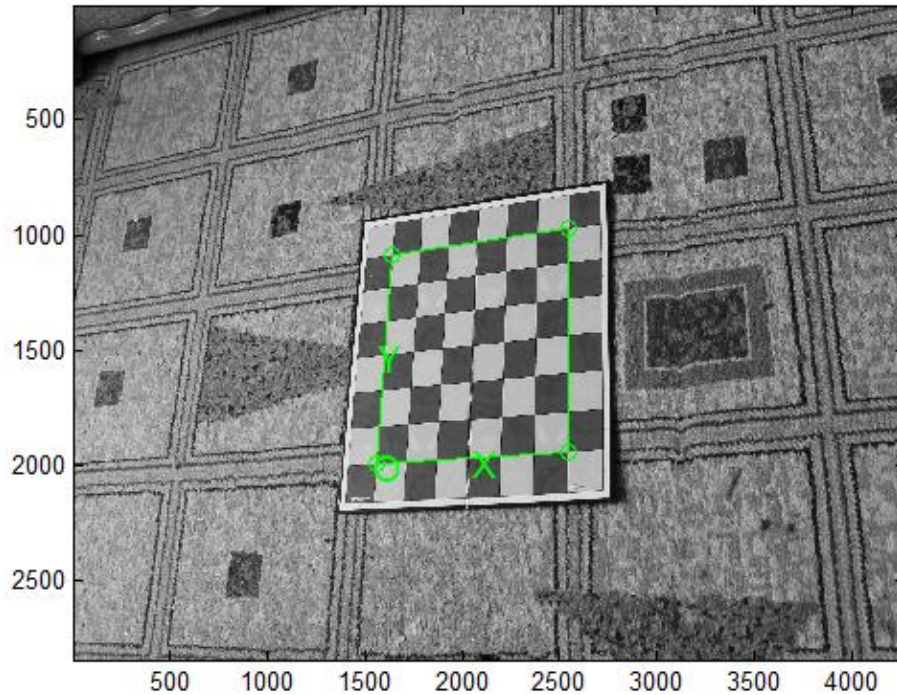
Fig. 33 Coordinates

Let **P** be a point space of coordinate vector **XX = [X;Y;Z]** in the grid reference frame (reference frame shown on the previous figure).

Let **XX$_c$ = [X$_c$;Y$_c$;Z$_c$]** be the coordinate vector of **P** in the camera reference frame.

Then **XX** and **XX$_c$** are related to each other through the following rigid motion equation (11):

$$XXc = Rc\_1 * XX + Tc\_1 \tag{11}$$

In particular, the translation vector **Tc_1** is the coordinate vector of the origin of the grid pattern (**O**) in the camera reference frame, and the thrid column of the matrix **Rc_1** is the surface normal vector of the plane containing the planar grid in the camera reference frame.

The same relation holds for the remaining extrinsic parameters (**Rc_2,Tc_2**), (**Rc_3,Tc_3**), ... , (**Rc_20,Tc_20**).

Once the coordinates of a point is expressed in the camera reference frame, it may be projected on the image plane using the intrinsic camera parameters.

65

The vectors **omc_1, omc_1, ... , omc_20** are the rotation vectors associated to the rotation matrices **Rc_1, Rc_1, ... , Rc_20**. The two are related through the rodrigues formula. For example, **Rc_1 = rodrigues(omc_1)**.

Similarly to the intrinsic parameters, the uncertainties attached to the estimates of the extrinsic parameters **omc_i, Tc_i** (**i=1,...,n_ima**) are also computed by the toolbox. Those uncertainties are stored in the vectors **omc_error_1,..., omc_error_20, Tc_error_1,..., Tc_error_20** (assuming **n_ima = 20**) and represent approximately three times the standard deviations of the errors of estimation.

Finding the dependence of the coordinates.

To define location and orientation of object on the second image, we need to construct dependence between object coordinates and the initial coordinate system of the second image. Further dependence between initial coordinates systems of both images. And dependence of the initial coordinate system of the first image to coordinate system of the camera.

We will assume that on the second image there is an object of P, which has been rotated around the x-axis α1 degrees; about the y-axis β1 degrees; and around the z-axis γ1 degrees, has coordinates (12):

$$P = \begin{pmatrix} 1 & 0 & 0 & x \\ 0 & 1 & 0 & y \\ 0 & 0 & 1 & z \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{12}$$

The dependence between object coordinates and the initial coordinate system of the second image (CS2 – initial coordinate system of the 2nd image) will be described on the following formula (13):

$$P2 = P \rightarrow CS2 \tag{13}$$

$$= P * \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & cos\alpha1 & sin\alpha1 & 0 \\ 0 & -sin\alpha1 & cos\alpha1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} * \begin{pmatrix} cos\beta1 & 0 & sin\beta1 & 0 \\ 0 & 1 & 0 & 0 \\ -sin\beta1 & 0 & cos\beta1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$* \begin{pmatrix} cos\gamma1 & sin\gamma1 & 0 & 0 \\ -sin\gamma1 & cos\gamma1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where the first matrix – rotation around x-axis; the second matrix – rotation around y-axis; the third matrix – rotation around z-axis; and the last matrix – translation.

After that we need to find dependence between initial coordinates systems of both images (CS1 – initial coordinate system of the 1st image) (14):

$$P1 = P2 \rightarrow CS1 \tag{14}$$

$$= P2 * \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & cos\alpha2 & sin\alpha2 & 0 \\ 0 & -sin\alpha2 & cos\alpha2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} * \begin{pmatrix} cos\beta2 & 0 & sin\beta2 & 0 \\ 0 & 1 & 0 & 0 \\ -sin\beta2 & 0 & cos\beta2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$* \begin{pmatrix} cos\gamma2 & sin\gamma2 & 0 & 0 \\ -sin\gamma2 & cos\gamma2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where the first matrix – rotation around x-axis $\alpha2$ degree; the second matrix – rotation around y-axis $\beta2$ degree; the third matrix – rotation around z-axis $\gamma2$ degree.

Next we need to find dependence between initial coordinate system of the first image to coordinate system of the camera (CScam – coordinate system of the camera) (15):

$$Pcam = P1 * CScam \tag{15}$$

$$= P1 * \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & cos\alpha3 & sin\alpha3 & 0 \\ 0 & -sin\alpha3 & cos\alpha3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} * \begin{pmatrix} cos\beta3 & 0 & sin\beta3 & 0 \\ 0 & 1 & 0 & 0 \\ -sin\beta3 & 0 & cos\beta3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$* \begin{pmatrix} cos\gamma3 & sin\gamma3 & 0 & 0 \\ -sin\gamma3 & cos\gamma3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} xc \\ yc \\ zc \\ 1 \end{pmatrix}$$

where the first matrix − rotation around x-axis α3 degree; the second matrix − rotation around y-axis β3 degree; the third matrix − rotation around z-axis γ3 degree; and the last matrix − translation with xc, yc, zc − coordinates of the camera.

So we can find the location and orientation of the object while moving the camera (Fig.34).
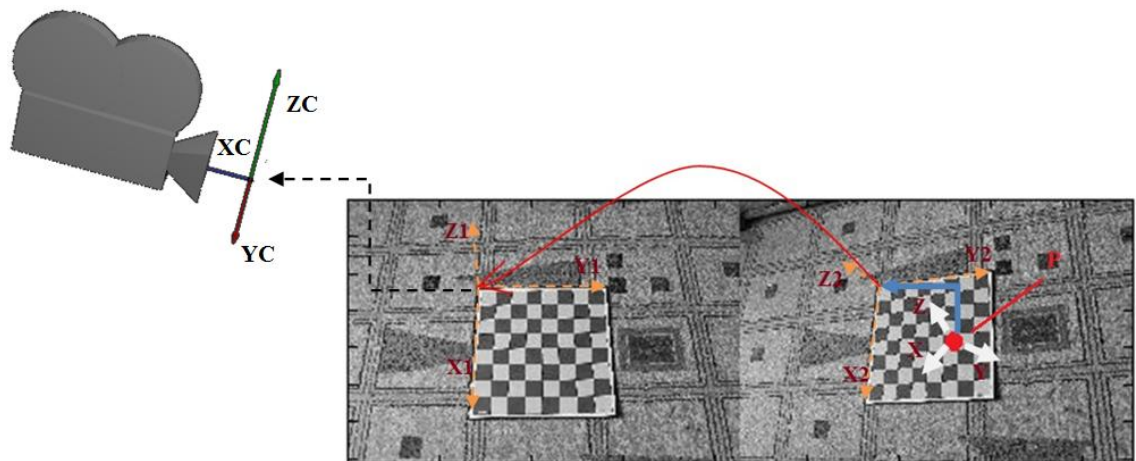

Fig. 34 Dependence of coordinates

On the previous figure you can see translation and rotation of the object's coordinates.

6 Conclusion

During development of a method the following tasks were solved: calibration of cameras; finding of relationship between images; finding of a repozition of the image; an object arrangement when moving the camera.

Also during work we reviewed history of computer vision, progress since the early work, problems and examples with computer vision. In thesis we considered various methods, sensors and systems the locations allowing to define and/or orientation of object. We investigated various descriptors and detectors which can be helpful in investigating computer vision. Was calibrated camera with finding its internal parameters using camera calibration toolbox for matlab. We found a relationship between coordinate systems and developed a method for determining the location and orientation of an object in space by using a video sensor and matlab.

7 Future work

In further research and development will be possible to use this method to obtain information about the location of the object. This can be used for entertainment purposes. Thus it will be possible not only to try conventional cameras for experimentation, but Kinect, Playstation chamber, and so on. Also using knowledge and abilities of programming, it is possible to develop this direction towards development of commercial programs, as example for purchases and sales of various large-size objects.

References

1. Donley EA. Nuclear Magnetic Resonance Gyroscopes, National Institute of Standards and Technology, 2010.

2. Sagnac G. L'ether lumineux demontre par l'effet du vent relatif d'ether dans un interferometre en rotation uniforme. Comptes Rendus, 1913.

3. Д. Л.Л., Лифшиц Е.М. Теория поля // Теоретическая физика, Vol. 2, 2006. P. 534.

4. В. С.И. Курс общей физики. 1st ed. Москва. 2004. 255-256 pp.

5. С. П.Д. Гироскопические приборы и системы. 2nd ed. Москва: Гироскопические системы. Ч. 2, 1988. 424 pp.

6. Дудник П.И., Чересов Ю.И. Авиационные радиолокационные устройства // ВВИА им. проф. Н.Е. Жуковского. 1986. P. 538.

7. Васин В.В. Радиолокационные устройства (теория и принципы построения). Москва: Сов. Радио, 1970.

8. Кондратенков Г.С., Фролов. А.Ю. Радиовидение. РЛС дистанционного зондирования земли. Москва: Радиотехника, 2005.

9. Pike J. GPS III Operational Control Segment (OCX).

10. GPS control segment map.

11. GPS Time Series.

12. Спутниковая суперсистема создается армией США // Военное обозрение.

13. Д.А.Тучин. Кодовые измерения псевдодальности системы GPS. Модель ошибок и априорная оценка точности определения вектора положения, Институт прикладной математики им. М.В. Келдыша, Москва, 2002.

14. В. В.П., В. М.А., and В. П.В. Повышение точности корректируемой иррациональной навигационной системы // Приборостроение, No. 12, 2014.

15. iPi Soft.

16. Organic Motion.

17. Stanford Markerless Motion Capture Project.

18. Интервью с директором компании iPi Soft - Михаилом Никоновым —

Компьютерная графика и анимация — Render.ru.

19. B. H., B. S. Determing Optical Flow // MIT Artificial Intelligence Laboratory. 1980. No. 572.

20. Z P. Rough Sets – Theoretical Aspects of Reasoning about Data // Kluwer Academic Publishers.

21. JR Q. C4.5: Programs for Machine Learning // Morgan Kaufmann Publishers.

22. T M. Machine Learning. McGraw Hill, 1997.

23. J. S., C T. Good features to track // IEEE. 1994. pp. 593-600.

24. D. A., M. N., and L. S. Robust Vehicle Detection through Multidimensional Classification for On Broad Video Based Systems // IEEE. 2007.

25. P. F., H. J., and S. S. Real-time tracking and outlier rejection with changes in illumination // ICCV'01. 2001. Vol. 1. pp. 684-689.

26. C. P., T P. A trainable system for object detection // Computer Vision. 2000. Vol. 38. No. 1. pp. 15-33.

27. Tuytelaars T., Mikolajczyk K. Local Invariant Feature Detectors: A Survey. 2008.

28. Zisserman, Fergus, and Perona. A sparse object category model for efficient learning and exhaustive recognition. In IEEE Computer Society Conference onComputer Vision and Pattern Recognition (CVPR'2005).

29. Z. K., J. M., and K. M. Forward-backward error: automatic detection of tracking failures. 2010. 2756-2759 pp.

30. Волосов Д.С. Оптические аберрации объективов // Фотографическая оптика. 1978. Vol. 2. pp. 91-543.

31. // Camera Calibration Toolbox for Matlab: [сайт].

32. Конушин А. Геометрические свойства нескольких изображений // Компьютерная графика и мультимедиа , No. 4(3), 2006.

33. Fishler M.A., Boles R.C. Random sample concensus: A paradigm for model fitting with applications to image analysis and automated cartography // Comm. Assoc. Comp. 1981. Vol. 24. No. 6. pp. 381-395.

34. Hartley R., Zisserman A. Multiple View Geometry in Computer Vision // Cambridge University Press. 2001. pp. 101-113.

35. Brown D.C. Decentering Distortion of Lenses // Photometric Engineering, Vol. 32, No. 3, 1966. pp. 444-462.

36. А. Ф. Всегда по курсу // Братишка : Ежемесячный журнал подразделений специального назначения., No. 4, 2013. pp. 22-25.

Appendix A   Matlab code

Thesis.m

clear

%Reading images and describing corner features detected in the image

im1 = iread('I1.jpg', 'mono', 'double');

im2 = iread('I2.jpg', 'mono', 'double');

h1 = icorner(im1, 'nfeat', 500, 'color', 'patch', 20)

h2 = icorner(im2, 'nfeat', 500, 'color', 'patch', 20)

%Match h1 and h2; display images

m = h1.match(h2)

m(1:5)

idisp({im1, im2})

m.subset(40).plot('g')

%Construct a histogram of error

[m,corresp] = h1.match(h2);

corresp(:,1:5)

m2 = h1.match(h2);

ihist(m2.distance, 'normcdf')

%Translate the matrix and display approximate position of the camera

T1 = transl(-0.1, 0, 0) * troty(0.4);

cam1 = CentralCamera('name', 'camera 1', 'default', ...

0.002, 'pose', T1)

T2 = transl(0.1, 0,0)*troty(-0.4);

cam2 = CentralCamera('name', 'camera 2', 'default', ...

0.002, 'pose', T2);

axis([-0.5 0.5 -0.5 0.5 0 1])

cam1.plot_camera('color', 'b', 'label')

cam2.plot_camera('color', 'r', 'label')

P=[0.5 0.1 0.8]';

 plot_sphere(P, 0.03, 'b');

%Plots the epipolar lines due to the fundamental matrix F and the image points p

p1 = cam1.plot(P)

```
p2 = cam2.plot(P)
cam1.hold
e1 = cam1.plot( cam2.centre, 'Marker', 'd', 'MarkerFaceColor', 'k')
cam2.hold
e2 = cam2.plot( cam1.centre, 'Marker', 'd', 'MarkerFaceColor', 'k')
F = cam1.F( cam2 )
rank(F)
null(F)'
e1 = h2e(ans)'
null(F');
e2 = h2e(ans)'
cam2.plot_epiline(F, p1, 'r')
cam1.plot_epiline(F', p2, 'r');

% And through all points
E = cam1.E(F)
sol = cam1.invE(E)
inv(cam1.T) * cam2.T
Q = [0 0 10]';
cam1.move(sol(:,:,1)).project(Q)
cam1.move(sol(:,:,2)).project(Q)
sol = cam1.invE(E, Q)
P = homtrans( transl(-1, -1, 2), 2*rand(3,20) );
p1 = cam1.project(P);
p2 = cam2.project(P);
F = fmatrix(p1, p2)
rank(F)
cam2.plot(P);
cam2.plot_epiline(F, p1, 'r')

% Find the smallest distance of point from epipolar line
p2(:,[8 7]) = p2(:,[7 8]);
fmatrix(p1, p2)
epidist(F, p1(:,1), p2(:,1))
```

```matlab
epidist(F, p1(:,7), p2(:,7))

[F,in,r] = ransac(@fmatrix, [p1; p2], 1e-6, 'verbose')

in

F = m.ransac(@fmatrix, 1e-4, 'verbose')

m.show

m(1:5)

idisp({im1, im2});

m.inlier.subset(100).plot('g')


%Display error detection point matches

idisp({im1, im2});

m.outlier.subset(100).plot('r')


% Homogeneous to Euclidean

cam = CentralCamera('image', im1);

cam.plot_epiline(F', m.inlier.subset(20).p2, 'g');

h2e( null(F))

cam.plot(ans, 'bo')


% Create grid of points

Tgrid = transl(0,0,1)*trotx(0.1)*troty(0.2);

P = mkgrid(3, 1.0, 'T', Tgrid);

p1 = cam1.plot(P, 'o');

p2 = cam2.plot(P, 'o');


% Build matrix of homography; display cam2

H = homography(p1, p2)

p2b = homtrans(H, p1);

cam2.hold()

cam2.plot(p2b, '+')


% Display the position of the camera relative frame

p1b = homtrans(inv(H), p2);

axis([-1 1 -1 1 0 2])

plot_sphere(P, 0.05, 'b')

plot_sphere(Q, 0.05, 'r')
```

```
cam1.plot_camera('color', 'b', 'label')
cam2.plot_camera('color', 'r', 'label')
```