Image-based navigation

Master's Degree in Geomatics Engineering and Geoinformation

Academic Year 2017-2018

E.T.S.I.Geodesica, Cartografíca y Topográfica

Image-based navigation

- ✓ Introduction
- ✓ Impacts on image-based navigation
- ✓ Image sequence analysis
- ✓ Image-based navigation techniques

✓ Based on Hofmann-Wellenhof *et al.*

Introducción

Image-based navigation aims at navigating objects by processing series of image data.

These image data may be recorded with passive sensors like digital cameras or active instruments like laser scanners.

Image-based navigation allows to extend the definition of navigation beyond the so far primarily geometric task: e.g. autonomous vehicles to find their way even within unknown surroundings.

Current fields of application often refer to robot technology, covering the operation and control of industrial robot arms up to the guidance of mobile robots.

Image-based navigation serves as one component of a multisensor navigation system and is responsible for specialized tasks like collision avoidance.

Image-based navigation has several roots: photogrammetry, digital image processing, and computer vision.

Definitions

- ✓ Single images

Single images are individual two- or three-dimensional projections of parts of the world which are of minor importance for navigation purposes. A single (or mono) image does not provide sufficient information for the determination or interpretation of motions within the scene.

- ✓ Simultaneous-image pairs

Simultaneous-image pairs arise from spatially separated projections of the object space, recorded by two or more prevalently convergent sensors at the same time. If two sensors are involved, the term stereo image pair is used.

Pairs of two-dimensional images allow for deducing three-dimensional geometric properties of the object space at a single epoch.

Definitions

✓ Image sequences

This term denotes series of images sequentially recorded by specific sensors at discrete time epochs.

The analysis of image sequences forms the backbone of all image-based navigation techniques.

Three-dimensional motions within the scene may be deduced. In the literature, tight and loose sequences are distinguished, reflecting the time interval between subsequent projections. Basically, geometric variations will be larger within loose sequences than within tight sequences.

In case that only a series of individual images is used, the term mono sequence applies. Otherwise, hybrid sequences are obtained where simultaneous-image pairs are available at every recording epoch of the sequence.

Navigation techniques

- ✓ Self positioning

  the sensor (or a set of sensors) is mounted on the moving object whose trajectory has to be determined

  appropriate for mobile robots or autonomous vehicles

- ✓ Remote positioning

  the sensor set remains stationary to track one oreven several remote objects

  used for stationary industrial applications

Navigation techniques

✓ Advantages

Detailed information about the environment of a moving object

Autonomous navigation technique which imitates the primary sense of orientation and navigation of human beings

Apart from positioning, it allows for attitude determination, object recognition, and active techniques like collision avoidance

✓ Disadvantages

Passive sensors → only directional views of the object space are available, which are further affected by occlusions

In unknown environments image-based self-positioning becomes a relative navigation technique → loss of accuracy over the distance traveled.

Self-positioning is not possible in "featureless" environments (visual navigation fails without landmarks)

Specific reference frames

✓ Sensor frame

Cartesian coordinate frame

The origin is the center of projection of the sensor

Terrestrial (close-range) photogrammetry:

x2 represents the optical axis of the sensor and points from the center of projection towards the object space ("forward")
x3 points to the zenith
x1~ completes the axes to a right-handed

Restricting to optical instruments, the sensor frame is complemented by the two-dimensional image frame defined within the image plane π of an optical sensor

$\mathbf{x}_1^p$ (parallel to $\mathbf{x}_1^s$) $\mathbf{x}_3^p$ (parallel to $\mathbf{x}_3^s$)

For the sake of simplicity, the image coordinates will be denoted by x and z

Specific reference frames

✓ Sensor frame

Interior orientation of the camera (relation between the image and sensor frame)

principal distance d $\rightarrow$ orthogonal distance of the center of projection from the image plane

coordinates xo, zo of the principal point of the image $\rightarrow$ point of intersection of the optical axis x2 with the image plane

Note that for cameras focused to infinity, the principal distance is equal to the focal length f. For the close-range cameras used in image-based navigation, d is greater than f and depends on the focus setting.

The principal point of the image usually differs from the origin of the image frame.

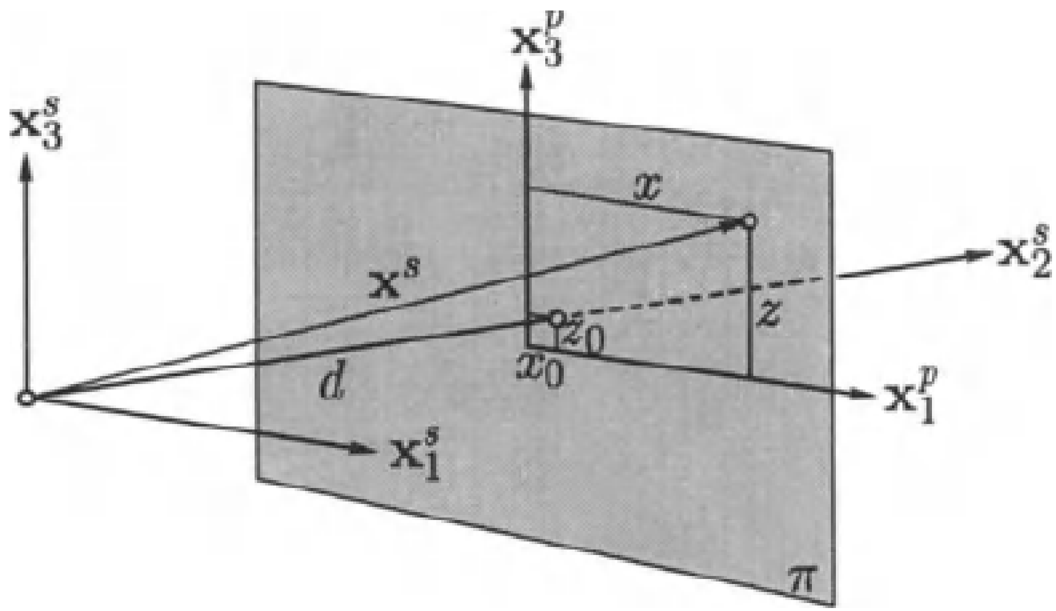
When transforming image coordinates to vectors in the sensor frame, the image data have to be centered at the principal point of the image

Thus, the point (x, z) in the image plane is given by the image vector

$$\mathbf{x}^s = \begin{bmatrix} x_1^s \\ x_2^s \\ x_3^s \end{bmatrix} = \begin{bmatrix} x - x_0 \\ d \\ z - z_0 \end{bmatrix}$$

Specific reference frames

- ✓ Sensor frame



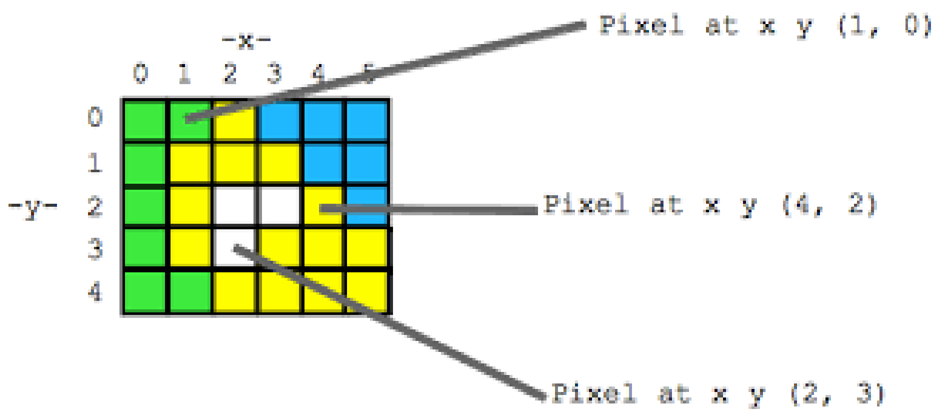
$$\mathbf{x}^s = \begin{bmatrix} x_1^s \\ x_2^s \\ x_3^s \end{bmatrix} = \begin{bmatrix} x - x_0 \\ d \\ z - z_0 \end{bmatrix}$$

Specific reference frames

✓ Sensor frame

Digital image frame coordinates must be considered when dealing with digital imagery

These are usually represented by a matrix of discrete picture elements (pixels) and the origin is the top left matrix element



This step is omitted here since the transformation between the (conventional) image frame and the digital image frame is a straightforward two-dimensional operation (translation of origin, mirroring of the z-axis)

Specific reference frames

✓ Model frame

This type of frame is only relevant for stereophotogrammetry

Local three-dimensional Cartesian coordinate frame resulting from the photogrammetric evaluation of a stereo pair of two-dimensional images (directional bundles)

The coordinate axes of the model frame are denoted by $\mathbf{x}_1^m,\ \mathbf{x}_2^m,\ \mathbf{x}_3^m$

When creating a three-dimensional model from two or more directional bundles, scale, position, and attitude of the model remain unknown

The transformation from the model frame to the *e-frame*, i.e., the earth-centered-earth-fixed (ECEF) frame

$$\mathbf{x}^e = \mathbf{x}^e_{0,m} + \mu\,\mathbf{R}^e_m(\alpha_1, \alpha_2, \alpha_3)\,\mathbf{x}^m$$

scale → scalar quantity μ
position → three translational components contained in the vector $\mathbf{x}_0^e$
attitude → defined by three rotation angles $\alpha_1,\ \alpha_2,\ \alpha_3$

Specific reference frames

✓ Model frame

The transformation from the model frame to the *e-frame*, i.e., the earth-centered-earth-fixed (ECEF) frame is defined by seven parametres

scale $\rightarrow$ scalar quantity μ

position $\rightarrow$ three translational components contained in the vector $\mathbf{x}_0^e$

attitude $\rightarrow$ defined by three rotation angles $\alpha_1,\ \alpha_2,\ \alpha_3$

The complete seven-parametres transformation (absolute orientation) is given by

$$\mathbf{x}^e = \mathbf{x}_{0,m}^e + \mu\,\mathbf{R}_m^e(\alpha_1, \alpha_2, \alpha_3)\,\mathbf{x}^m$$

Sensors

✓ Passive and active sensors

passive sensors → digital cameras

- only record the radiation reflected or emitted by the environment
- yield directional information in planar images referred to an image frame
- the inverse problem of photogrammetry has to be solved (derivation of three-dimensional properties from two-dimensional images)
- the complete sensitive area is usually exposed simultaneously

active sensors → laser scanners

- provide artificial radiation and record the returned ener
- primarily provide distance information (travel time of energy)
- directional information may be obtained by a continuou realignment of the radiation beam
- Increased recording intervals (compared to passive)

Passive sensors

- ✓ In high-precision photogrammetry, high-end analog measurement cameras are still in use due to their superior technical features (concerning geometric image quality, resolution, and other features).
- ✓ However, modern digital image sensor technology has led to an increased performance of digital sensors in view of the spatial and radiometric quality of the image data.
- ✓ Creating an image with a digital optical sensor involves four quantization steps:

    1. A single image is only a "snapshot" of a small part of the world, which corresponds to a temporal quantization of the four-dimensional spacetime continuum
    2. The optical projection restricts the three-dimensional space to the twodimensional image plane
    3. The image plane itself is partitioned into a limited number of lightsensitive pixels with finite dimensions resulting in a spatial quantization of the image contents
    4. The final conversion from analog to digital image data denoted as analog-to-digital (AD) conversion implies a radiometric quantization of the image data. The "real" intensities present in the object space are transformed to discrete intensity levels of the digital image.

Passive sensors

- ✓ The temporal, spatial, and spectral resolution of the sensor should be tuned to the planned application. For many tasks, monochrome sensors are sufficient.
- ✓ The exposure interval of the sensor should be kept as low as possible to reduce the problem of motion blurring (Le., image distortions due to motions occurring during the exposure interval)
- ✓ Current digital image sensors are based on arrays of solid-state semiconductor elements where incoming photons are converted to charge pairs taking advantage of the photoelectric effect.
- ✓ Charge-coupled-device (CCD) technology has dominated digital imaging since the early 1980s.
- ✓ More recently, active pixel sensors (APS) relying on complementary metal-oxide semiconductors (CMOS) have become a valuable alternative and are already replacing CCD sensors in many applications.
- ✓ The APS technology provides random access to the pixels and allows integrating electronic functionality or even complete digital processors within the image sensors. Thus, such sensors may be actively tuned to the application needs leading to the notation "smart sensors"

Passive sensors

- ✓ Among other features, digital image sensors are characterized by the number of pixels, the physical dimensions of the pixels, the spectral sensitivity of the semiconductor elements, the quantum efficiency (i.e., the percentage of incident photons actually converted to charge pairs and successfully detected), the signal-to-noise ratio (SNR), and the dynamic range of the sensor pixels

- ✓ Another important property of optical sensors is the depth of field (or depth of focus), representing the range between the closest and farthermost point (with respect to the center of projection) along the optical axis that are projected with acceptable blurring. Apart from the object distance, the depth of field is determined by the lens system and aperture of the sensor

Active sensors

✓ Laser

- Lasers operate in or close to the visible band of the electromagnetic spectrum with typical wavelengths between 550 nm (blue-green laser) and 1540 nm (infrared laser)
- For three-dimensional laser scanning, the alignment of the laser beam is controlled by optical components or deflecting elements. The scanning technique is designated as lidar (light detection and ranging)
- The strength of the reflected signal may be used to generate intensity images
- Due to the precise alignment control and the small laser apertures, lidar has become a highly accurate remote sensing technique with a gradually increasing number of applications

✓ Radar

The day and night capability as well as the weather independence of radar are advantages compared to passive imaging or laser. Typically, two-dimensional radar images involving horizontal direction and distance are used. The intensity of the pixels depends on the strength of the back scatter. The main drawbacks of radar are the low geometric resolution of the images due to the large aperture of the radar radiation cone and the relatively long scanning times. The geometric resolution may be improved by synthetic aperture radar (SAR)

Impacts on image-based navigation

✓ Photogrammetry

The geometric fundamentals of image-based navigation are closely related to photogrammetry

These techniques are extensively used in cartography, earth observation, and geoinformation disciplines.

In an increasing number of applications, traditional photographic techniques are being replaced by digital sensing and processing (both passive and active).

Another important issue is mobile mapping, a close-range technique.

The main tasks of photogrammetry are related to sensor orientation,i.e., the determination of the position and attitude of the sensor (or set of sensors), and the determination of the coordinates of points in the object space.

The techniques applied vary between mono and stereo (or multi-) image approaches.

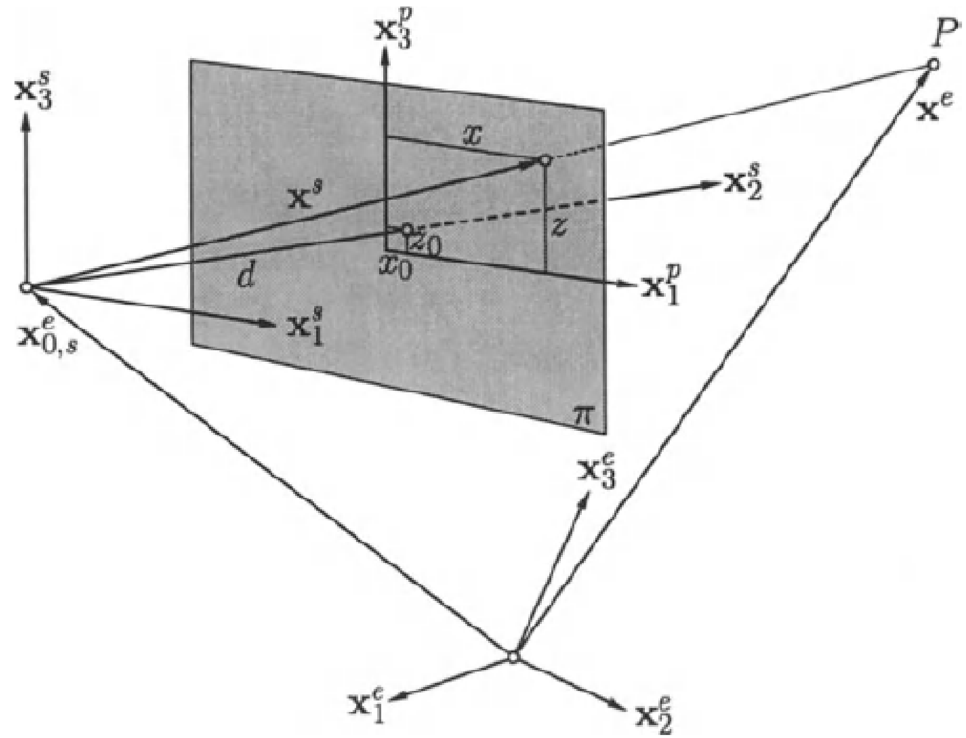
The transformations are needed since the measurement data are gained within the sensor (s) frame, whereas the point coordinates (e.g., of a moving object or vehicle) are usually desired in the ECEF (e) frame.

Impacts on image-based navigation

- ✓ Perspective projection

$$\nu\,\mathbf{x}^s = \mathbf{R}_e^s\,(\mathbf{x}^e - \mathbf{x}_{0,s}^e)$$

$$\mathbf{x}^e = \mathbf{x}_{0,s}^e + \nu\,\mathbf{R}_s^e\,\mathbf{x}^s$$



Impacts on image-based navigation

✓ Interior orientation

The interior (or inner) orientation of an optical sensor describes the current position of the center of projection with respect to the image plane.

The corresponding calibration parameters are *d,* the principal distance of the sensor, and *xo, xo,* the coordinates of the principal point in the image.

In many cases, interior orientation also aims at calibrating the optical distortion of the lens system of the camera.

In contrast to the high-end sensors of aerial photogrammetry, the parameters of interior orientation are usually not constant for the digital sensors used in image-based navigation.

One way of determining the calibration parameters is to set up a network of control points whose e-frame coordinates are known and to perform the calibration with high redundancy. In this case, interior and exterior orientation (see below) of the sensor are usually determined together by applying a technique known as bundle adjustment. To guarantee a high-quality calibration,
a proper spatial distribution of the control points is required.

Impacts on image-based navigation

- ✓ Monoimage techniques

The position and attitude of an imaging sensor is denoted as exterior (or outer) orientation.

The determination of exterior orientation leads to the transformation between the *s-frame* and *e-frame*.

The transformation parameters include three translations represented by the vector $\mathbf{x}_0^e$ and three orientation angles contained in the rotation matrix $\mathbf{R}_e^s$

Geometrically, the determination of exterior orientation corresponds to spatial resection.

The solution is usually performed iteratively starting with an approximate solution

Impacts on image-based navigation

- ✓ Stereo-image techniques

Stereoimaging represents the classical approach to circumvent the inverse problem of photogrammetry by introducing a second image of the scene obtained from a different viewpoint

The derivation of spatial information from stereo images requires that the two images overlap

The measurement of parallaxes is the basic concept of stereo imaging

Impacts on image-based navigation

- ✓ Stereo-image techniques

A parallax corresponds to the apparent shift of the position of an object in the image plane caused by a change of the position of the observer

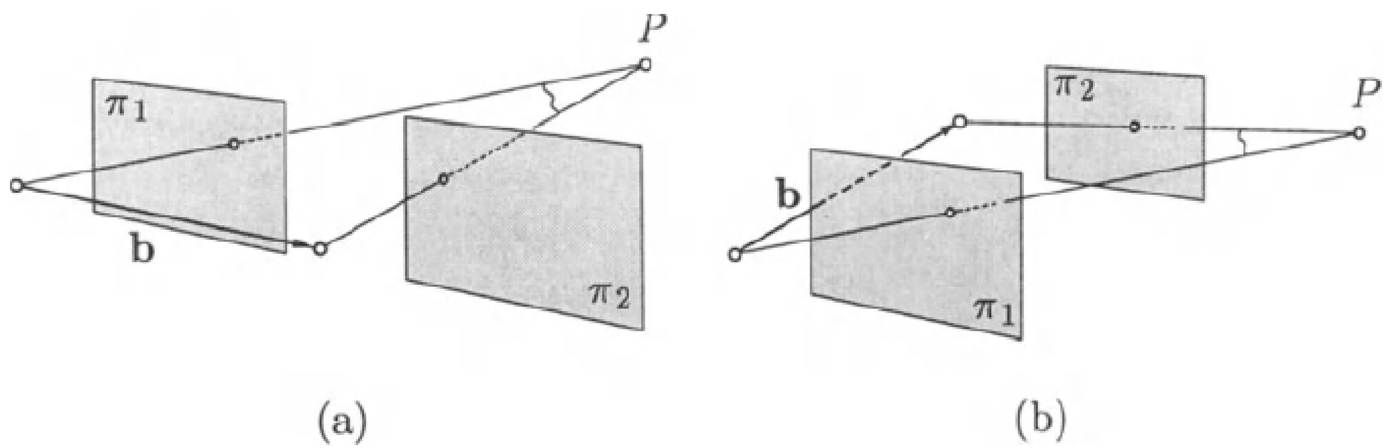


Fig. 12.3. Standard (a) and axial (b) setup in stereophotogrammetry

Impacts on image-based navigation

- ✓ Relative orientation

The determination of the relative orientation of an image pair is based on the use of homologous image points, i.e., projections of identical object points contained in both images

Considering the reconstruction relation and introducing the indices for the two sensor frames, the reconstruction of a homologous point P from both images is given by

$$\mathbf{x}^e = \mathbf{x}^e_{0,s_1} + \nu_{s_1} \mathbf{R}^e_{s_1} \mathbf{x}^{s_1} ,$$

$$\mathbf{x}^e = \mathbf{x}^e_{0,s_2} + \nu_{s_2} \mathbf{R}^e_{s_2} \mathbf{x}^{s_2} ,$$

Impacts on image-based navigation

✓ Relative orientation

When differencing the previous Eqs.   cancels, which reflects that the coordinates of the homologous points need not be known in the *e-frame*.

A slight rearrangement yields

$$\mathbf{x}^e_{0,s_2} - \mathbf{x}^e_{0,s_1} = \nu_{s_1} \mathbf{R}^e_{s_1} \mathbf{x}^{s_1} - \nu_{s_2} \mathbf{R}^e_{s_2} \mathbf{x}^{s_2}$$

$$\mathbf{b}^e = \mathbf{x}^e_{0,s_2} - \mathbf{x}^e_{0,s_1} \qquad \mathbf{b}^e = \nu_{s_1} \mathbf{R}^e_{s_1} \mathbf{x}^{s_1} - \nu_{s_2} \mathbf{R}^e_{s_2} \mathbf{x}^{s_2}$$

$$(\mathbf{R}^e_{s_2} \mathbf{x}^{s_2}) \cdot \left(\mathbf{b}^e \times (\mathbf{R}^e_{s_1} \mathbf{x}^{s_1})\right) = 0$$

Note that this Eq. still involves nine unknowns

Impacts on image-based navigation

✓ Relative orientation

The desired number of only five unknowns for relative orientation is achieved by considering only relative orientation angles between the two images and eliminating the scale of the model by fixing the length of the baseline vector to an arbitrary value.

Although a minimal number of five homologous points are required, relative orientation will usually be performed with high redundancy yielding an improved quality of the orientation parameters.

Assuming that the relative orientation of the image pair is known, the condition of coplanarity may be reused to find homologous points in the two images which allow to generate a three-dimensional model of the object space (reconstruction).

If a point P is given in one image, the corresponding point in the second image lies on a certain line denoted as epipolar line.

Thus, the homologous point may be identified by a simplified search along the epipolar line.

Impacts on image-based navigation

✓ Absolute orientation

- The result of relative orientation is a three-dimensional model of the object
- space given in the *m-frame*.

- The task of absolute orientation is the transformation of this model into the e-frame, requiring a seven-parameter transformation.

- The solution requires the use of known control points.

- Since three-dimensional positions of the control points are given in the e and m-frames, every control point contributes three equations for the determination of absolute orientation.

- Thus, the theoretical minimum of control points is two plus one component of a third control point. The use of redundant control points is recommended.

Impacts on image-based navigation

✓ Determination of object points

- When using a single image, additional information about the object space is required to resolve the inverse problem of photogrammetry.
- Thus, monoimage techniques may be used when a detailed model of the object space is available.
- In an absolutely oriented stereo image pair, the determination of object points is performed by trigonometric methods. Note that the coordinates of a homologous point given in both images lead to a redundant pointpositioning task: each pair of image coordinates determines a ray given in the three-dimensional e-frame that originates from the corresponding center of projection and points towards the image point.
- Due to unavoidable measurement errors, the two rays will not intersect. This problem may be overcome by least squares adjustment.

Digital image processing

✓ Navigation requires digital image processing

Since most navigation applications require online information about the state vector of a moving object or vehicle, traditional optical sensors as well as traditional analysis methods are unsuitable for image-based navigation.

Instead, digital sensors and digital image processing techniques are required.

✓ Motivations for processing digital images

- image improvement
- feature extraction
- accentuation of certain contents
- visualization of image data

In principle, only image improvement and feature extraction are relevant for image-based navigation. Both techniques are based on image filtering.

Digital image processing

- ✓ Image filtering

Filters are classified into low-pass, high-pass, and band-pass filters.

Another classification is made between space-domain and frequency-domain filters.

The latter use simpler filtering algorithms but require a preceding transformation
of the image data into the frequency domain, which is a time consuming procedure.

Thus, space-domain filters are usually preferred in image-based navigation.

Digital image processing

✓ Space-domain filtering as a general technique

Discrete two-dimensional convolution of the digital image represented by the intensity (or gray level) matrix G with a filter matrix F whose layout is defined by the filter type

$$\bar{G}(x,z) = \sum_{k=-m}^{k=+m} \sum_{l=-m}^{l=+m} G(x+k, z+l)\, F(m+k, m+l)$$

$\bar{G}(x,z)$ represents the filter result

$G(x,z)$ represents the original intensity at the central pixel

$F(k,l)$ element (k,l) of the filter matrix $(2m+1, 2m+1)$

Digital image processing

- ✓ Image improvement

In certain cases, the quality of digital images may be insufficient, limiting the extraction potential of valuable information from the image data.

Image corrections may be utilized for repairing disturbing effects of the atmosphere or of the sensor type used, and for overcoming possible quality limitations caused by directional-sensitive reflections within the object space.

These techniques are commonly referred to as noise suppression.

Noisy image data often contain "disturbing" pixels characterized by a significant difference in intensity compared to the surrounding pixels.

Such disturbances may affect or even prevent the extraction of characteristic features from the data.

To overcome this problem, the image data may be manipulated by applying low-pass filtering. These filters emphasize low-frequency information and suppress high-frequency information. The moving-average
and the Gaussian filter belong to the most common low-pass filters.

Digital image processing

✓ Image improvement

As an example, a moving-average filter is shown where the eight surrounding pixels of the central pixel are considered. The filter matrix readsIn certain cases, the quality of digital images may be insufficient, limiting the extraction potential of valuable information from the image data.

$$\mathbf{F} = \frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Since the intensities of the pixels are added to determine the filtered intensity of the central pixel, low-pass filters are also denoted as additive filters.

Gaussian filters are used to artificially reduce the resolution of an image (subsampling). The reason for this process may be the reduction of storage space or a technique denoted as depth from focus, i.e., the determination of the depth of a point on the basis of the focus setting of the sensor and the blurring level of the image.

Digital image processing

✓ Feature extraction

Contrary to image improvement, feature extraction (or image segmentation) is based on high-pass filtering of the image data, i.e., high-frequency information is emphasized, whereas low-frequency information is suppressed.

Therefore, feature extraction is sensitive to high-frequency image disturbances.

In digital photogrammetry and, thus, in image-based navigation, the automatic extraction of characteristic image features is of major interest for the relative orientation of images.

Note that the identification of corresponding features in different images is crucial for the quality of relative orientation. Problems of automatic feature extraction arise in featureless image regions; however, human operators would encounter the same difficulties when dealing with such data.

Digital image processing

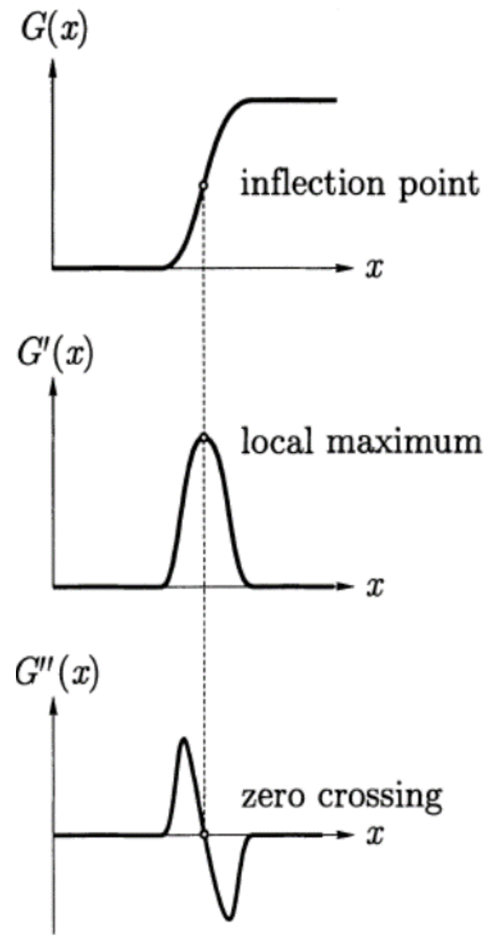
✓ Feature extraction

A characteristic feature within an image is usually associated with a significant change of intensity.

One method to determine the position of this change of intensity is based on forming derivatives of the intensity function into different directions

As an example for high-pass filtering, the Sobel operator for the x direction is shown, which is a gradient filter

$$\mathbf{F}_x = \frac{1}{4}\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$



Digital image processing

✓ Feature extraction

Characteristic image features to be extracted are mainly points, edges, or regions.

Only characteristic points are discussed here.

Several criteria define a characteristic point:

- **significance**: the point differs significantly from its surrounding
- **rarity**: the point is dissimilar to others
- **interpretability**: the point supports the interpretation of the image
- **invariance**: the point is invariant with respect to geometric and temporal changes
- **stability**: the point is resistant against image disturbances
- **quality**: the position of the point may be determined without biases

Note that some of these criteria are in competition with each other: while low-pass filtering may contribute to the stability of the point, it may also reduce its significance and quality.

Computer vision

The third impact on image-based navigation is computer vision.

Although the distinction between digital image processing and computer vision is not clearly defined, the latter is rather seen as the development of computer based techniques for imitating the way human beings perceive (i.e., see and interpret) the world with their eyes.

Computer vision, an important discipline of artificial intelligence (AI), may be divided into three major research fields:

machine vision → deals with close-range industrial applications with controlled illumination and relatively simple environments

image understanding → aims at constructing a geometric model of the scene with simple features ("primitives") for image segmentation and object recognition

active vision → deals with tasks like collision avoidance or target tracking

Computer vision

- ✓ Object recognition as an example of computer vision

Object recognition is required for tasks like obstacle detection or robot control.

For the purpose of collision avoidance, the detection of objects that might prevent planned maneuvers is usually sufficient.

For certain robot applications, however, it may be necessary to recognize a family of objects or even to identify individual representatives within this family of objects.

Thus, the general term of recognition involves three levels of detail

Detection

Recognition

Identification

Computer vision

✓ Object recognition as an example of computer vision

Object recognition has to address three fundamental questions:

1. What are appropriate processing techniques for extracting features from the image data that are related to a certain object?

2. What is a proper way of modeling an object to enable a comparison of the model with those structures extracted from the image data?

3. What are possible strategies for establishing valid correspondence relations between extracted features and existing object models (matching)?

The first question is a problem of image processing.

The second and the third question deal with the required strategies for representing and identifying objects within knowledge databases.

Various approaches are proposed in the literature.

Computer vision

✓ Object recognition as an example of computer vision

Some exemplary properties of object models are listed below

- Surface models vs. volume models: object models may be based on geometric descriptions of the object surface (e.g., surfaces of generalized cylinders or cones); or they may consist of a set of volume elements (voxels) yielding a discrete approximation of the object

- Explicit or exemplary storage vs. relational structures: object models may comprise a large number of directional views of a certain object; or they could be based on geometric primitives (like spheres or cubes) that are connected to each other following certain topological rules like "above", "to the right", etc. While exemplary storage is only suitable for simple objects, relational structures are also applicable to deformable objects

- Geometric vs. radiometric models: some objects may be classified by their geometric properties, whereas others may have specific radiometric features due to unique reflectance patterns of the object surface

Computer vision

✓ Object recognition as an example of computer vision

It cannot be generally stated that a certain type of model is ideal. The type to be chosen strongly depends on the application and on the type of object to be represented. Note that some objects are exactly defined by their geometry (such as industrial construction elements), whereas others have a generic geometry, meaning that the basic structures are identical but every individual object has its own specific features.

The complexity of object recognition in real-world image sequences arises from the polymorphism of natural objects. In many cases, it is not sufficient to set up object models on the basis of stationary knowledge databases but learning algorithms are required for scene interpretation purposes.

The recognition of an object may follow a data-driven approach by analyzing the image data and trying to identify objects included in the data, or it may follow a top-down strategy by searching a given object within the image data. The latter procedure is usually applied for identification purposes.

One important problem of object recognition is that only directional views of the objects exist. (Further, occlusions may hide parts of the object).

Image sequence analysis

✓ General aspects

The central element of image-based navigation is the image sequence recorded by a sensor (or a set of sensors).

Applying appropriate analysis techniques, geometric motion information may be deduced from the image data, which finally enables to solve the navigation task.

Image sequence analysis supports a very heterogeneous field of applications including diagnostic medicine, industrial quality control, meteorology, etc.

The major aim of image sequence analysis is to establish correspondences between subsequent images of a sequence, and, thus, to deduce geometric relationships between the contents of the images.

Essentially, this means interpreting intensity variations within the image plane as geometric changes motion estimation is an exclusively geometric task which does not involve an interpretation (e.g., motions) within the scene. Note that of the observed scene.

Image sequence analysis

✓ Analysis criteria

Motion estimation and interpretation are influenced by a number of criteria characterizing the image sequence. The most fundamental impact arises from possible sensor motions

Remote positioning → a set of usually stationary sensors track the motions of one or several objects.

Self-positioning → the sensor (set) itself is in motion. Image sequences recorded with moving sensors are more difficult to analyze since all contents of the image plane fluctuate

Motions occurring within the scene are of similar significance. Generally, the interpretation of stationary scenes is less complicated, especially if this property is known a priori.

In case of nonstationary scenes, typically, one or several objects move in front of a stationary background. In case of self-positioning, possible knowledge of the structure of the scene may reduce the online computational burden.

Image sequence analysis

✓ Analysis criteria

With respect to the character of the sensor and/or scene motions, their magnitude, type (i.e., translation and/or rotation), and degree of freedom influence the interpretation of the image data.

Translations are usually tracked more easily.

Object rotations may support object recognition by relieving the problem of directional views.

In case of remote positioning, the physical properties of the objects to be tracked should be taken into account, e.g., whether they are rigid or deformable. The physical properties will influence the type of object model to be used if object recognition is required.

Image sequence analysis

✓ Analysis criteria

Finally, the type and number of sensors is a decisive factor.

With laser scanners, three-dimensional information about the scene is available from a single image.

Using passive sensors, spatial information must be deduced from multiple images (i.e., simultaneous pairs of images or sequential images of a sequence).

If the geometric properties of the scene are known, monoimage techniques may be applied.

Summarizing, the most challenging task is the analysis of monoimage sequences recorded with a moving passive sensor in an unknown environment!

This scenario will be assumed for the further discussion of image sequence analysis.

Image sequence analysis

✓ Processing strategies

Tasks of image sequence analysis:

1. Image correspondence $\rightarrow$ establishment of correspondences and optical flow

2. Derivation of projected motion $\rightarrow$ to deduce projected motion from optical flow

3. Reconstruction of motion and structure $\rightarrow$ involves the derivation of the actual three-dimensional sensor and/or scene motions from the projected motion

Image sequence analysis

✓ Image correspondence

Establishment of correspondences between subsequent images of the sequence by the determination of optical flow

A single image of the object space yields the intensity matrix for a discrete projection epoch. Taking another image at another epoch (forming a motion stereo), the contents of the image plane and, thus, the intensity matrix will change

This temporal variation of the intensity matrix is denoted as optical flow, and it is usually described by means of a discrete vector field.

Several sources may cause a variation of the intensity matrix between subsequent images

- sensor motion
- motions within the scene
- changes of illumination
- changes of surface reflections

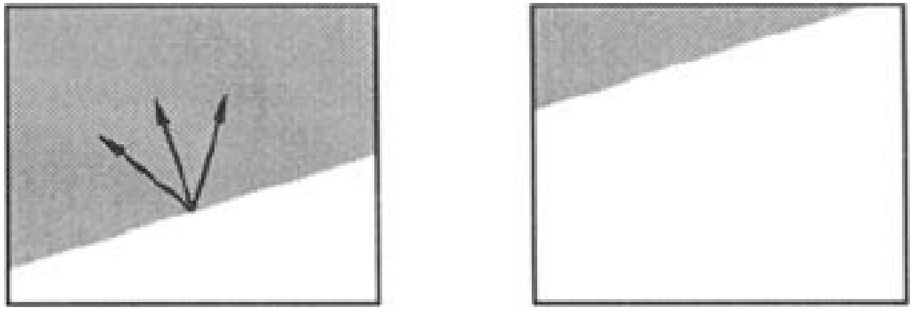
Image sequence analysis

✓ Image correspondence

When navigating, both sensor motion and motions within the scene are to be deduced from the image sequence.

In contrast, changes of illumination and reflection are experienced as disturbances.

Another difficulty of image correspondence is the aperture problem → motion interpretation cannot be done uniquely (asindicated by the arrows)



Only displacements orthogonal to the projected edge may be detected

Image sequence analysis

✓ Derivation of projected motion

The designation projected motion relates to the perspective projection of real motions into the image plane

The mathematical model is derived from the analytical form of perspective projection

$$\nu\,\mathbf{x}^s = \mathbf{R}_e^s\,(\mathbf{x}^e - \mathbf{x}_{0,s}^e) \qquad (*)$$

In case of a moving sensor, the six parameters of exterior orientation (included in $\mathbf{R}_e^s$ and $\mathbf{x}_0^e$) and the image coordinates x, z of all projected points become time-dependent

The temporal changes of x and z are found by forming the time derivative of (*) yielding the velocities

$$u = dx/dt \qquad w = dz/dt$$

These velocities are used to express the projected motion and are deduced from optical flow

Image sequence analysis

- ✓ Derivation of projected motion

The determination of projected motion requires the elimination of disturbing effects due to illumination and reflection variations

In a stationary scene, the direction and structure of the optical-flow vector field allows for determining the type of motion exerted by the sensor (i.e., translation and/ or rotation)

Object boundaries typically yield discontinuities of the optical flow since close objects tend to show a faster motion in the image plane than distant objects, which is due to perspective projection

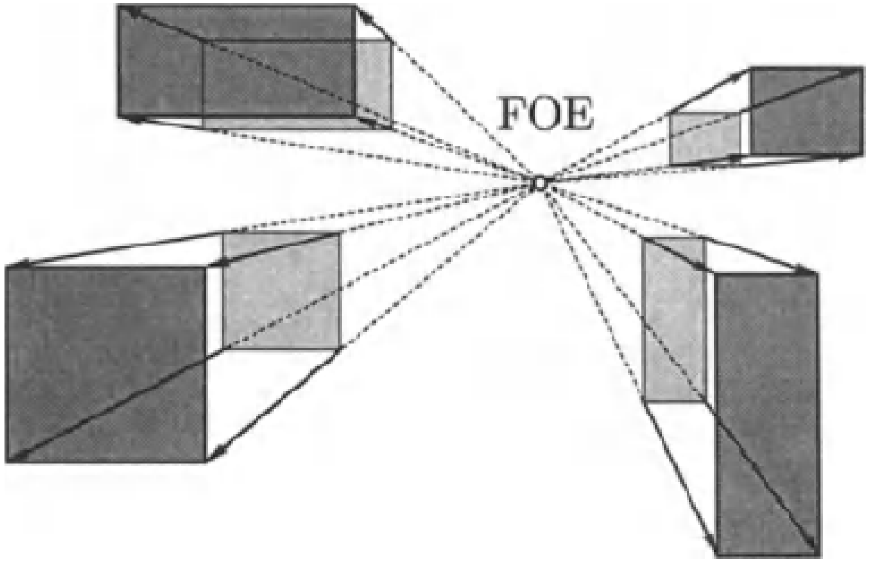
Image sequence analysis

- ✓ Derivation of projected motion

In case of translation, the resulting vector field shows a radial structure

All vectors seem to arise from a common origin that indicates the vanishing point of all lines parallel to the current direction of motion

At this point, the optical flow is zero



All contents of the image plane diverge radially from that point (in case of forward motion), which is therefore called focus of expansion (FOE)

In case of backward motion, the term focus of contraction (FOC) is used

Image sequence analysis

- ✓ Derivation of projected motion

For a known FOE (or FOC), the distance of each object point from the center of projection is indirectly proportional to its divergence speed, i.e., the higher the divergence speed the closer the point

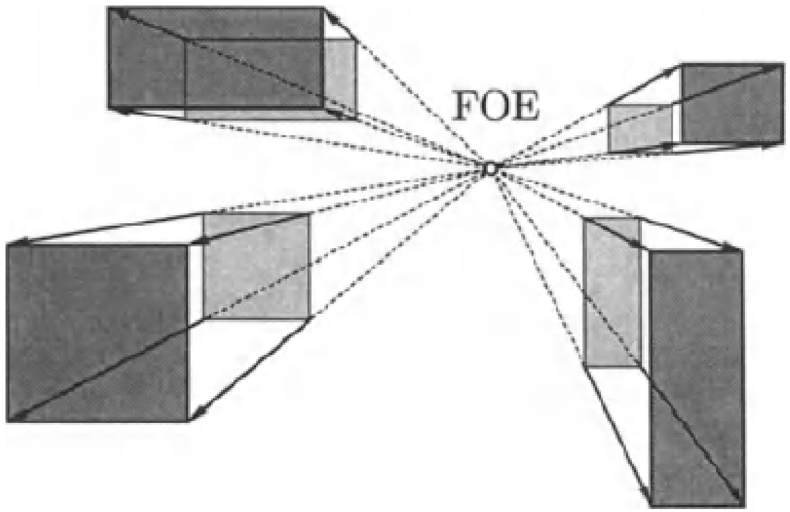


Fig. 12.6. Focus of expansion and optical flow of planar objects in case of translation

If the optical flow vectors of at least two (stationary) points are known, the FOE may be derived as the point of intersection between these vectors

The determination of the FOE is a valuable step for image correspondence since it may support the detection of homologous points or regions within subsequent images (like the epipolar lines)

Image sequence analysis

- ✓ Reconstruction of motion and structure
- ✓ To determine the actual three-dimensional sensor and/or scene motions from the projected motion
- ✓ This task is closely related to the inverse problem of photogrammetry, i.e., the derivation of three-dimensional information from two-dimensional imagery
- ✓ A "byproduct" of image-based navigation is a three-dimensional model of the object space that is implicitly generated in the context of image sequence analysis (reconstruction of structure).
- ✓ From a photogrammetric point of view, image sequence analysis means repeated relative and absolute orientation of new images with respect to previous sections of the sequence.

Image sequence analysis

- ✓ Reconstruction of motion and structure
- ✓ Reconstruction of motion and structure may be based on:
- ✓ Inversion approaches → three-dimensional information about the scene has to be derived from the image data. These techniques are also known as shape from motion or shape from stereo.
- ✓ Projection approaches → based on an existing model of the scene which may have different levels of detail. The model is artificially projected into the image plane, and correspondence is found by comparing these data with the real projection obtained by the sensor

Image sequence analysis

- Suummary of the main elements of image sequence generation and analysis

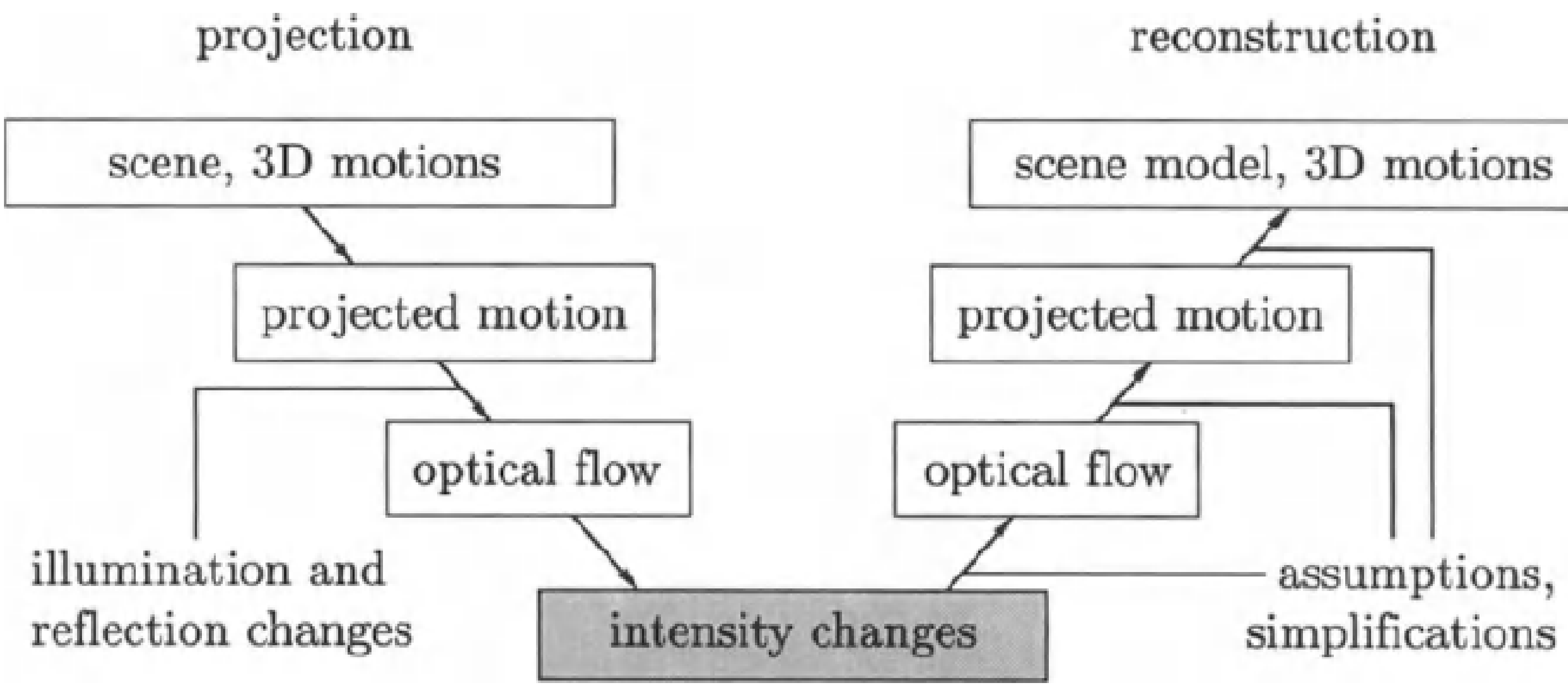


Image-based navigation techniques

Since the actual requirements of navigation tasks may differ significantly, only a brief overview of image-based navigation techniques is given for

- ✓ Self positioning
- ✓ Remote positioning

Image-based navigation techniques

✓ Self-positioning

The main tasks within image-based self-positioning may be classified into two groups.

1. Preparatory steps (preparatory steps)

- sensor calibration → requires an accurately surveyed network of control points
- initial exterior orientation → since the technique is relative, an initial navigation state vector is required

2. Recursive steps (to be repeatedly performed during the nav.pro)

- image correspondence
- determination of projected motion
- reconstruction of three-dimensional motion and structure
- modeling of the sensor trajectory

Image-based navigation techniques

✓ Recursive steps

The main tasks within image-based self-positioning may be classified into two groups.

1. Preparatory steps (preparatory steps)

- sensor calibration → requires an accurately surveyed network of control points
- initial exterior orientation → since the technique is relative, an initial navigation state vector is required

2. Recursive steps (to be repeatedly performed during the nav.pro)

- image correspondence
- determination of projected motion
- reconstruction of three-dimensional motion and structure
- modeling of the sensor trajectory

Image-based navigation techniques

- ✓ Example of recursive sensor orientation

Note that the coordinates of the points A, B, C are determined as a by product

Using an oriented image, five homologous points are required in the next image of the scene to determine the relative orientation of the new image relative to the previous one

In contrast, the exterior orientation of a single image requires a minimum of three known scene points

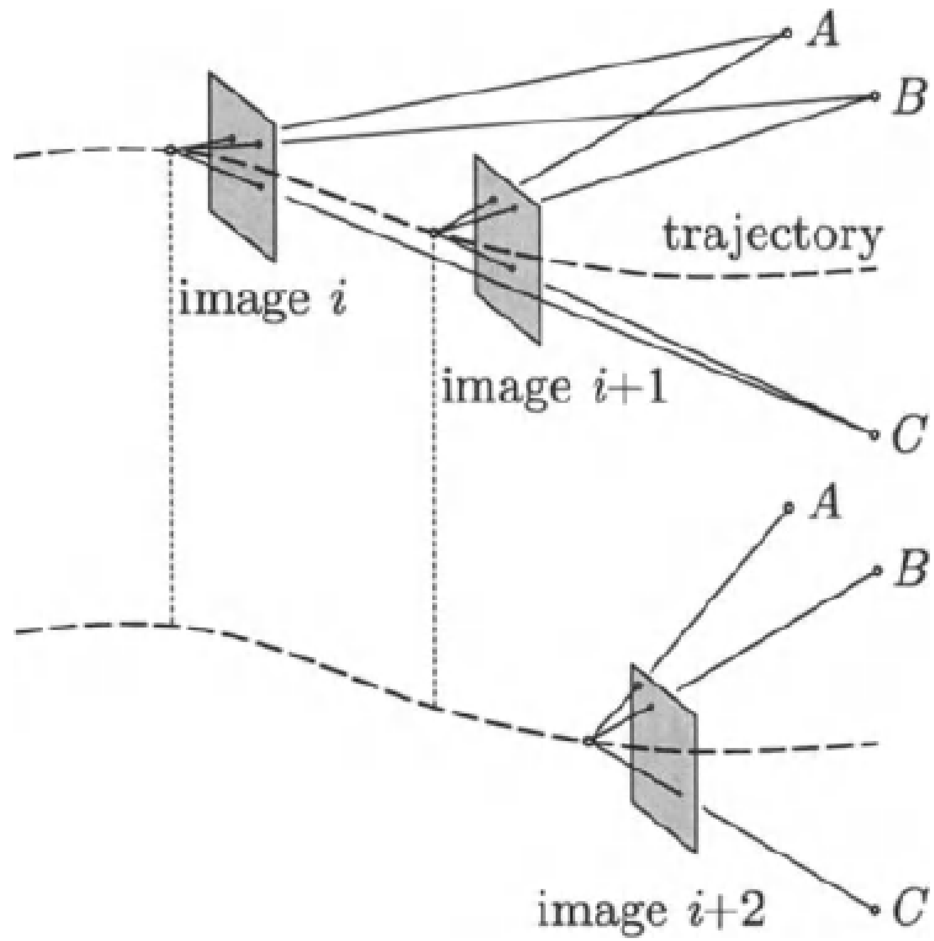


Image-based navigation techniques

- ✓ Remote positioning

In image-based remote positioning, a set of at least two sensors is used to track the motions of one or several remote objects.

Considering only a single epoch, two or more simultaneous images with known relative and absolute orientation are available.

In principle, the tasks to be solved resemble those of self-positioning.

The main differences are that image correspondence is strongly simplified due to the availability of simultaneous and relatively oriented images, and that it is not necessary to model the trajectory of the sensors.

The major drawback of remote positioning is its limitation to a locally bounded surrounding, which is defined by the layout of the sensor setup.