**Methods for registration of graphical items in images:** Image registration is the process of overlaying two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors. It geometrically aligns two images—the reference and sensed images. The present differences between images are introduced due to different imaging conditions. Image registration is a crucial step in all image analysis tasks in which the final information is gained from the combination of various data sources like in image fusion, change detection, and multichannel image restoration.

In general, its applications can be divided into four main groups according to the manner of the image acquisition:

Different viewpoints (multiview analysis). Images of the same scene are acquired from different viewpoints. The aim is to gain larger a 2D view or a 3D representation of the scanned scene. Examples of applications: Remote sensing—mosaicing of images of the surveyed area. Computer vision—shape recovery (shape from stereo).

Different times (multitemporal analysis). Images of the same scene are acquired at different times, often on regular basis, and possibly under different conditions. The aim is to find and evaluate changes in the scene which appeared between the consecutive image acquisitions. Examples of applications: Remote sensing—monitoring of global land usage, landscape planning. Computer vision—automatic change detection for security monitoring, motion tracking. Medical imaging—monitoring of the healing therapy, monitoring of the tumor evolution.

Different sensors (multimodal analysis). Images of the same scene are acquired by different sensors. The aim is to integrate the information obtained from different source streams to gain more complex and detailed scenerepresentation.

Examples of applications: Remote sensing—fusion of information from sensors with different characteristics like panchromatic images, offering better spatial resolution, color/multispectral images with better spectral resolution, or radar images independent of cloud cover and solar illumination. Medical imaging—combination of sensors recording the anatomical body structure like magnetic resonance image (MRI), ultrasound or CT with sensors monitoring functional and metabolic body activities like positron emission tomography (PET), single photon emission computed tomography (SPECT) or magnetic resonance spectroscopy (MRS). Results can be applied, for instance, in radiotherapy and nuclear medicine.

Scene to model registration. Images of a scene and a model of the scene are registered. The model can be a computer representation of the scene, for instance maps or digital elevation models (DEM) in GIS, another scene with similar content (another patient), ‘average’ specimen, etc. The aim is to localize the acquired image in the scene/model and/or to compare them.

Examples of applications: Remote sensing—registration of aerial or satellite data into maps or other GIS layers.

Nevertheless, the majority of the registration methods consists of the following four steps:

* Feature detection. Salient and distinctive objects (closed-boundary regions, edges, contours, line intersections, corners, etc.) are manually or, preferably, automatically detected. For further processing, these features can be represented by their point representatives (centers of gravity, line endings, distinctive points), which are called control points (CPs) in the literature.
* Feature matching. In this step, the correspondence between the features detected in the sensed image and those detected in the reference image is established. Various feature descriptors and similarity measures along with spatial relationships among the features are used for that purpose.
* Transform model estimation. The type and parameters of the so-called mapping functions, aligning the sensed image with the reference image, are estimated. The parameters of the mapping functions are computed by means of the established feature correspondence.
* Image resampling and transformation. The sensed image is transformed by means of the mapping functions. Image values in non-integer coordinates are computed by the appropriate interpolation technique.

**Hough transform for line-cuts:** A straight line is defined by two points A = (x1, y1) and B = (x2, y2). All the lines going through point **A** are given by the expression y1 = kx1 + q, this same rule applies also for straight lines passing through point B y2 = kx2 + q. The only common point of both straight lines in the k, q parameter space is the point which in the original image space represents the only existing straight line connecting points A and B. This means that any straight line in the image is represented by a single point in the k,q parameter space and any part of this straight line is transformed into the same point. The main idea of line detection is to determine all the possible line pixels in the image, to transform all lines that can go through these pixels into corresponding points in the parameter space, and to detect the points (a, b) in the parameter space that frequently resulted from the Hough transform of lines y = ax + b in the image. Detection of all possible line pixels in the image may be achieved by applying an edge detector to the image; then, all pixels with edge magnitude exceeding some threshold can be considered possible line pixels. In the most general case, nothing is known about lines in the image, arid therefore lines of any direction may go through any of the edge pixels. In reality, the number of these lines is infinite; however, for practical purposes, only a limited number of line directions may be considered. The possible directions of lines define a discretization of the parameter k. Similarly, the parameter q is sampled into a limited number of values. The parameter space is not continuous any more, but rather is represented by a rectangular structure of cells. This array of cells is called the accumulator array A, whose elements are accumulator cells A(k, q).

For each edge pixel, parameters k, q are determined which represent lines of allowed directions going through this pixel. For each such line, the values of line parameters *k,* q are used to increase the value of the accumulator cell A(k, q). Clearly, if a line represented by an equation ***y = ax+ b*** is present in the image, the value of the accumulator cell A(a, b) will be increased many times-as many times as the line ***y = ax + b*** is detected as a line possibly going through any of the edge pixels. For any pixel P, lines going through it may have any direction k (from the set of allowed directions), but the second parameter q is constrained by the image co-ordinates of the pixel *P* and the direction ***k.*** Therefore, lines existing in the image will cause large values of the appropriate accumulator cells in the image, while other lines possibly going through edge pixels, which do not correspond to lines existing in the image, have different k, q parameters for each edge pixel, and therefore the corresponding accumulator cells arc increased only rarely. In other words, lines existing in the image may be detected as high-valued accumulator cells in the accumulator array, and the parameters of the detected line are specified by the accumulator array co-ordinates. As a result, line detection in the image is transformed to detection of local maxima in the accumulator space.

It has been noted that an important property of the Hough transform is its insensitivity to missing parts of lines, to image noise, and to other non-line structures co-existing in the image.

Note that the parametric equation of the line ***y = kx + q*** is appropriate only for explanation of the Hough transform principles-it causes difficulties in vertical line detection (k -> ∞) and in non-linear discretization of the parameter k. If a line is represented **as: **

the Hough transform does not suffer from these limitations.

Discretization of the parameter space is an important part of this approach**;** also, detecting the local maxima in the accumulator array is a non-trivial problem. In reality, the resulting discrete parameter space usually has more than one local maximum per line existing in the image, and smoothing the discrete parameter space may be a solution. All these remarks remain valid if more complex curves are sought in the image using the Hough transform, the only difference being the dimensionality of the accumulator may.

**Algorithm 6.14: Curve detection using the Hough transform**

1. Quantize parameter space within the limits of parameters a. The dimensionality n of the parameter space is given by the number of parameters of the vector a.
2. Form an n-dimensional accumulator array A(a) with structure matching the quantization of parameter space; set all elements to zero.
3. For each image point (xl, ***x2)*** in the appropriately threshold gradient image, increase all accumulator cells A(a) if f (x, a) = 0: for all a inside the limits used in step 1.
4. Local maxima in the accumulator array A(a) correspond to realizations of curves f (x, a) that are present in the original image.

If we are looking for circles, the analytic expression ***f*** (x, a) of the desired curve is: where the circle ***has*** center (a, b) and radius r. Therefore, the accumulator data structure must be three-dimensional. For each pixel x whose edge magnitude exceeds a given threshold, all accumulator cells corresponding to potential circle centers (a, b) are incremented in step **3** of the given algorithm. The accumulator cell A(a, b, r) is incremented if the point *(a,* b) is at distance r from point **x:** and this condition is valid for all triplets (a, b, r) satisfying the equation. If some potential center (a, b) of a circle of radius *r* is frequently found in the parameter space, it is highly probable that a circle with radius *r* and center (a: b) really exists in the processed data. The processing results in a set of parameters of desired curves f (x, a) = 0 that correspond to local maxima of accumulator cells in the parameter space; these maxima best match the desired curves and processed data.

The desired region borders can rarely be described using a parametric boundary curve with a small number of parameters; in this case, a generalized Hough transform can offer the solution. This method constructs a parametric curve (region border) description based on sample situations detected in the learning stage. Assume that shape, size, and rotation of the desired region are known. A reference point xR is chosen at any locator inside the sample region, and then an arbitrary line can be constructed starting at this reference point aiming in the direction of the region border. The border direction (edge direction) is found at the intersection of the line and the region border. Areference table is constructed, and intersection parameters are stored as a function of the border direction at the intersection point; using different lines aimed from the reference point, all the distances of the reference point to region borders and the border directions at the intersections can be found.

**Algorithm 6.15: Generalized Hough transform**

1. Construct an R-table description of the desired object.
2. Form a data structure A that represents the potential reference points:



Set all accumulators cell values to zero.

1. For each pixel (x1, x2) in a thresholded gradient image, determine the edge direction Φ(x); find all potential reference points xR and increase all:



 for all possible values of rotation and size change:

 

1. The location of suitable regions is given by local maxima in the A data structure.

**Links to Radon transforms**: is the integral transform consisting of the integral of a function over straight lines. Radon further included formulas for the transform in three-dimensions, in which the integral is taken over planes. The Radon transform is widely applicable to tomography, the creation of an image from the scattering data associated to cross-sectional scans of an object. If a function ƒ represents an unknown density, then the Radon transform represents the scattering data obtained as the output of a tomographic scan. Hence the inverse of the Radon transform can be used to reconstruct the original density from the scattering data, and thus it forms the mathematical underpinning for tomographic reconstruction. The Radon transform data is often called a sinogram because the Radon transform of a Dirac delta function is a distribution supported on the graph of a sine wave. Consequently the Radon transform of a number of small objects appears graphically as a number of blurred sine waves with different amplitudes and phases.

Let *ƒ* be a continuous function vanishing outside some large disc in the Euclidean plane **R**2. The Radon transform, denoted by *Rƒ*, is a function defined on the space of lines *L* in **R**2 by



where the integration is performed with respect to the arc length measure *d*σ on *L*. Concretely, any straight line *L* can be parameterized by



where *s* is the distance of *L* from the origin and α is the angle *L* makes with the *x* axis. It follows that the quantities (α,*s*) can be considered as coordinates on the space of all lines in **R**2, and the Radon transform can be expressed in these coordinates by



More generally, in the *n*-dimensional Euclidean space **R***n*, the Radon transform of a compactly supported continuous function *ƒ* is a function *Rƒ* on the space Σ*n* of all [hyperplanes](http://en.wikipedia.org/wiki/Hyperplane) in **R***n*. It is defined by



for ξ ∈ Σ*n*, where the integral is taken with respect to the natural hypersurface measure *d*σ. Observe that any element of Σ*n* is characterized as the solution locus of an equation



where α ∈ *Sn*−1 is a unit vector and *s* ∈ **R**. Thus the *n*-dimensional Radon transform may be rewritten as a function on *Sn*−1×**R** via



It is also possible to generalize the Radon transform still further by integrating instead over *k*-dimensional affine subspaces of **R***n*. The X-ray transform is the most widely used special case of this construction, and is obtained by integrating over straight lines.

**Application for text & graphic recognition:**

Application for degraded text recognition:



Characters cannot be segmented by simple thresholding, and the color, size, font, and orientation of the text are unknown. The main design choice is the kind of text occurrences, between scene text and document text.

Text is considered as a scene text when the text is recorded from a part of a scene (e.g., road signs, posters on the street, street names). Unlike document text, characters in scene images originally exist in 3D space, and can therefore be distorted by a slant or a tilt, and by the shape of objects on which they are printed. Text extraction from a natural scene has been studied, in projects such as vehicle license plates detection.

**Texture segmentation:** Text detection technique is based on a texture segmentation approach. Text in a document is considered as a textured region to isolate; non-text contents in the image, such as blanks, pictures, graphics, and other objects in the image, must be considered as regions with different textures. The human vision can quickly identify text regions without having to recognize individual characters because text has textural properties that differentiate it from the rest of a scene. Instinctively, text has the following distinguishing characteristics.

1. Characters contrast with their background.
2. Text possesses some frequencies and orientation information.
3. Text shows spatial cohesion: characters appear in clusters at a regular distance aligned to a virtual line.

**Text characterization:** By processing text as a distinctive texture, we propose a text characterization based on a bank of Gabor filters associated with an edge density measure. The features are designed to identify text paragraphs. None of them will uniquely identify text regions. Each individual feature will still confuse text with non-text regions but a collection of features will complement each other and allow identifying text unambiguously. Physically interpreted, the Gabor transform acts like the Fourier transform but only for a small Gaussian window over the image. In spatial domain, the two-dimensional Gabor filter *h*(*x*, *y*) is given by:



where *σx* and *σy* are the standard deviations of the Gaussian envelope along the *x* and *y* directions, and *wx* and *wy* are the centered frequencies of the filter. One important characteristic of Gabor filter is its orientation selectivity, which can be understood when the expression of 2D Gabor filter is rewritten in polar coordinates as:





**Text region clustering:** use of an adapted *K*-means clustering algorithm to cluster feature vectors. In order to reduce computational time, we apply the standard *K*-means clustering to a reduced number of pixels and a minimum distance classification is used to categorize all surrounding non-clustered pixels. Empirically, the number of clusters (value of *K*) was set to three, value that works well with all test images. The cluster whose center is closest to the origin of feature vector space is labeled as background while the furthest one is labeled as text. Text boxes rotation is applied after the estimation of document skew. The angle is estimated from the shape and the centroids of all text boxes. The final stage of text detection module is a validation module that confirms text boxes. It identifies false text boxes by using heuristic rules about aspect ratio, global intensity indicators, and so forth. We have applied text detection module on a set of 100 test images where there are one or two text areas per image. A text zone is correct when all the lines of the text area are included. A partly detected text zone is considered as an error. On the contrary, a false detection occurs when the detected zone does not contain any text information.