1

Introduction

In the last few decades, the evolution in technology has provided a rapid development in image acquisition and processing, leading to a growing interest in related research topics and applications including image registration. Registration is defined as the estimation of a geometrical transformation that aligns points from one viewpoint of a scene with the corresponding points in the other viewpoint. Registration is essential in many applications such as video coding, tracking, detection and recognition of object and face, surveillance and satellite imaging, structure from motion, simultaneous localization and mapping, medical image analysis, activity recognition for entertainment, behaviour analysis and video restoration. It is considered one of the most complex and challenging problems in image analysis with no single registration algorithm to be suitable for all the related applications due to the extreme diversity and variety of scenes and scenarios. This book presents image, video and 3D data registration techniques for different applications discussing also the related quality performance metrics and datasets. State-of-the-art registration methods based on the targeted application are analysed, including an introduction to the problems and limitations of each method. Additionally, various assessment quality metrics for registration are presented indicating the differences among the related research areas. For example, the important features in a medical image (e.g. MRI data) may not be the same as in the picture of a human face, and therefore the quality metrics are adjusted accordingly. Therefore, state-of-the-art metrics for quality assessment are analysed explaining their advantages and disadvantages, and providing visual examples separately for each of the considered application areas.

1.1 The History of Image Registration

In image processing, one of the first times that the concept of registration appeared was in Roberts' work in 1963 [1]. He located and recognized predefined polyhedral objects

Image, Video & 3D Data Registration: Medical, Satellite & Video Processing Applications with Quality Metrics, First Edition. Vasileios Argyriou, Jesus Martinez Del Rincon, Barbara Villarini and Alexis Roche.

in scenes by aligning their edge projections with image projections. The first registration applied to an image was in the remote sensing literature. Using sum of absolute differences as similarity measure, Barnea and Silverman [2] and Anuta [3, 4] proposed some automatic methods to register satellite images. In the same years, Leese [5] and Pratt [6] proposed a similar approach using the cross-correlation coefficient as similarity measure. In the early 1980s, image registration was used in biomedical image analysis using data acquired from different scanners measuring anatomy. In 1973, for the first time Fischler and Elschlager [7] used non-rigid registration to locate deformable objects in images. Also, non-rigid registration was used to align deformed images and to recognize handwritten letters. In medical imaging, registration was employed to aligned magnetic resonance (MR) and computer tomography (CT) brain images trying to build an atlas [8, 9].

Over the last few years due to the advent of powerful and low-cost hardware, real-time registration algorithms have been introduced, improving significantly their performance and accuracy. Consequently, novel quality metrics were introduced to allow unbiased comparative studies. This book will provide an analysis of the most important registration methodologies and quality metrics, covering the most important research areas and applications. Through this book, all the registration approaches in different applications will be presented allowing the reader to get ideas supporting knowledge transfer from one application area to another.

1.2 Definition of Registration

During the last decades, automatic image registration became essential in many image processing applications due to the significant amount of acquired data. With the term *image registration*, we define the process of overlaying two or more images of the same scene captured in different times and viewpoints or sensors. It represents a geometrical transformation that aligns points of an object observed from a viewpoint with the corresponding points of the same or different object captured from another viewpoint. Image registration is an important part of many image processing tasks that require information and data captured from different sources, such as image fusion, change detection and multichannel image restoration. Image registration techniques are used in different contexts and types of applications. Typically, it is widely used in computer vision (e.g. target localization, automatic quality control), in remote sensing (e.g. monitoring of the environment, change detection, multispectral classification, image mosaicing, geographic systems, super-resolution), in medicine (e.g. combining CT or ultrasound with MR data in order to get more information, monitor the growth of tumours, verify or improve the treatments) and in cartography updating maps. Image registration is also employed in video coding in order to exploit the temporal relationship between successive frames (i.e. motion estimation techniques are used to remove temporal redundancy improving video compression and transmission).

In general, registration techniques can be divided into four main groups based on how the data have been acquired [10]:

- *Different viewpoints (multiview analysis)*: A scene is acquired from different viewpoints in order to obtain a larger/panoramic 2D view or a 3D representation of the observed scene.
- *Different times (multitemporal analysis)*: A scene is acquired in different times, usually on a regular basis, under different conditions, in order to evaluate changes among consecutive acquisitions.
- *Different sensors (multimodal analysis)*: A scene is acquired using different kinds of sensors. The aim is to integrate the information from different sources in order to reveal additional information and complex details of the scene.
- *Scene to model registration*: The image and the model of a scene are registered. The model can be a computer representation of the given scene, and the aim is to locate the acquired scene in the model or compare them.

It is not possible to define a universal method that can be applied to all registration tasks due to the diversity of the images and the different types of degradation and acquisition sources. Every method should take different aspects into account. However, in most of the cases, the registration methods consist of the following steps:

- *Feature detection*: Salient objects, such as close-boundary regions, edges, corners, lines and intersections, are manually or automatically detected. These features can be represented using points such as centre of gravity and line endings, which are called control points (CPs).
- *Feature matching*: The correspondence between the detected features and the reference features is estimated. In order to establish the matching, features, descriptors and similarity measures among spatial relationships are used.
- *Transform model estimation*: According to the matched features, parameters of mapping functions are computed. These parameters are used to align the sensed image with the reference image.
- *Image resampling and transformation*: The sensed image is transformed using the mapping functions. Appropriate interpolation techniques can be used in order to calculate image values in non-integer coordinates.

1.3 What is Motion Estimation

Video processing differs from image processing due to the fact that most of the observed objects in the scene are not static. Understanding how objects move helps to transmit, store and manipulate video in an efficient way. Motion estimation is the research area of imaging a video processing that deals with these problems, and it is also linked to feature matching stage of the registration algorithms. Motion

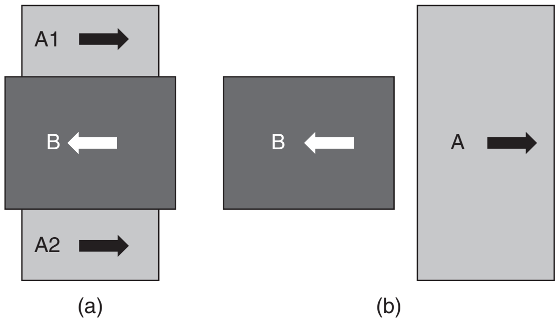


Figure 1.1 (a) Occluded objects A1 and A2, (b) single object A

estimation is the process by which the temporal relationship between two successive frames in a video sequence is determined. Motion estimation is a registration method used in video coding and other applications to exploit redundancy mainly in the temporal domain.

When an object in a 3D environment moves, the luminance of its projection in 2D is changing either due to non-uniform lighting or due to motion. Assuming uniform lighting, the changes can only be interpreted as movement. Under this assumption, the aim of motion estimation techniques is to accurately model the motion field. An efficient method can produce more accurate motion vectors, resulting in the removal of a higher degree of correlation.

Integer pixel registration may be adequate in many applications, but some problems require sub-pixel accuracy, either to improve the compression ratio or to provide a more precise representation of the actual scene motion. Despite the fact that sub-pixel motion estimation requires additional computational power and execution time, the obtained advantages settle its use that is essential for the most multimedia applications.

In a typical video sequence, there is no 3D information about the scene contents. The 2D projection approximating a 3D scene is known as 'homography', and the velocity of the 3D objects corresponds to the velocity of the luminance intensity on the 2D projection, known as 'optical flow'. Another term is 'motion field', a 2D matrix of motion vectors, corresponding to how each pixel or block of pixels moves. General 'motion field' is a set of motion vectors, and this term is related to the 'optical flow' term, with the latter being used to describe dense 'motion fields'.

Finding the motion between two successive frames of a video sequence is an ill-posed problem due to the intensity variations not exactly matching the motion of the objects. Another problematic phenomenon is the covered objects, in which case it is efficient to make the assumption that the occluded objects can be considered as many separable objects, until they are observed as a single object (Figure 1.1). Additionally, in motion estimation, it is assumed that motion within an object is smooth and uniform due to the spatial correlation.

The concept of motion estimation is used in many applications and is analysed in the following chapters providing details of state-of-the-art algorithms, allowing the reader to apply this information in different contexts.

1.4 Video Quality Assessment

The main target in the design of modern multimedia systems is to improve the video quality perceived by the user. Video quality assessment is a difficult task because many factors can interfere on the final result.

In order to obtain quality improvement, the availability of an objective quality metric that represents well the human perception is crucial. Many methods and measures have been proposed aiming to provide objective criteria that give accurate and repeatable results taking into account the subjective experience of a human observer. Objective quality assessment methods based on subjective measurements are using either a perceptual model of the human visual system (HVS) or a combination of relevant parameters tuned with subjective tests [11, 12].

Objective measurements are used in many image and video processing applications since they are easy to apply for comparative studies. One of the most popular metrics is peak signal-to-noise ratio (PSNR) that is based on the mean square error between the original and a distorted data. The computation of this value is trivial but has significant limitations. For example, it does not correlate well with the perceived quality, and in many cases the original undistorted data (e.g. images, videos) may not be available.

At the end of each chapter, a description of the metrics used to assess the quality of the presented registration methods is available for all the discussed applications, highlighting the key factors that affect the overall quality, the related problems and solutions, and the examples to illustrate these concepts.

1.5 Applications

1.5.1 Video Processing

Registration techniques are required in many applications based on video processing. As mentioned in the earlier section, motion estimation is a registration task employed to determine the temporal relationship between the video frames. One of the most important applications of motion estimation is in video coding systems.

Video CODECs (COder/DECoder) comprise an encoder and a decoder. The encoder compresses (encodes) video data resulting in a file that can be stored or streamed economically. The decoder decompresses (decodes) encoded video data (whether from a stored file or streamed), enabling video playback.

Compression is a reversible conversion of data to a format that requires fewer bits, usually performed so that the data can be stored or transmitted more efficiently. The size of the data in compressed form C relative to the original size O is known as the

Figure 1.2 Motion estimation predicts the contents of each macroblock base due the motion relative to the reference frame. The reference frame is searched to find the 16×16 block that matches the macroblock

compression ratio $R = O/C$. If the inverse of the process, 'decompression', produces an exact replica of the original data, then the compression is lossless. Lossy compression, usually applied to image and video data, does not allow reproduction of an exact replica of the original data but results in higher compression ratios.

Neighbouring pixels within an image or a video frame are highly correlated (spatial redundancy). Also neighbouring areas within successive video frames are highly correlated too (temporal redundancy).

A video signal consists of a sequence of images. Each image can be compressed individually without using the other video frames (intra-frame coding) or can exploit the temporal redundancy considering the similarity among consecutive frames (inter-frame coding), obtaining a better performance. This is achieved in two steps:

1. *Motion estimation*: A region (usually a block) of the current frame is compared with neighbouring region of the adjacent frames. The aim is to find the best match typically in the form of motion vectors (Figure 1.2).
2. *Motion compensation*: The matching region from the reference frame is subtracted from the current region block.

Motion estimation considers images of the same scene acquired in different time, and for this reason it is regarded as an image registration task. In Chapter 2, the most popular motion estimation methods for video coding are presented.

Motion estimation is not only utilised in video coding applications but also to improve the resolution and the quality of the video. If we have multiple, shifted and low-resolution images, we can use image processing methods in order to obtain high-resolution images.

Furthermore, digital videos acquired by consumer camcorders or high-speed cameras, which can be used in industrial applications and to track high-speed objects, are often degraded by linear space-varying blur and additive noise. The aim of video restoration is to estimate each image or frame, as it would appear without the effects of sensor and optics degradations. Image and video/restoration are essential when we want to extract still images from videos. This is because blurring and noise may not be visible to the human eye at usual frame rates, but they can become rather evident when observing a ‘freeze-frame’. The restoration is also a technique used when historical film materials are encoded in a digital format. Especially if they are encoded with block-based encoders, many artefacts may be present in the coded frame. These artefacts are removed using sophisticated techniques based on motion estimation. In Chapter 8, video registration techniques used in restoration applications are presented.

1.5.2 Medical Applications

Medical images are increasingly employed in health care for different kinds of tasks, such as diagnosis, planning, treatment, guided treatment and monitoring diseases progression. For all these studies, multiple images are acquired from subjects at different times and in the most of the cases using different imaging modalities and sensors. Especially with the growing number of imaging systems, different types of data are produced. In order to improve and gain information, proper integration of these data is highly desirable. Registration is then fundamental in this integration process. One example of different data registration is the epilepsy surgery. Usually the patients undergo various data acquisition processes including MR, CT, digital subtraction angiography (DSA), ictal and interictal single-photon emission computed tomography (SPECT) studies, magnetoencephalography (MEG), electroencephalography (EEG) and positron emission tomography (PET). Another example is the radiotherapy treatment, in which both CT and MR are employed. Therefore, it can be argued that the benefits for the surgeons are significant by registering all these data. Registering methods are also applied to monitor the growth of a tumour or to compare the patient’s data with anatomical atlases.

Motion estimation is also used for medical applications operating like a doctor’s assistant or guide. For example, motion estimation is used to indicate the right direction for the laser, displaying the optical flow (OF) during interstitial laser therapy (ILT) of a brain tumour. The predicted OF velocity vectors are superimposed on the grey-scaled images, and the vectors are used to predict the amount and the direction of heat deposition.

Another growing application of registration is in recognition of face, lips and feelings using motion estimation. A significant amount of effort has been put on sign language recognition. The motion and the position of the hand and the fingers are estimated, and patterns are used to recognize the words and the meanings (see Figure 1.3), an application particularly useful for deaf-mute people [13].

The main problems of medical image data analysis and the application of registration techniques are discussed in details in Chapter 7.

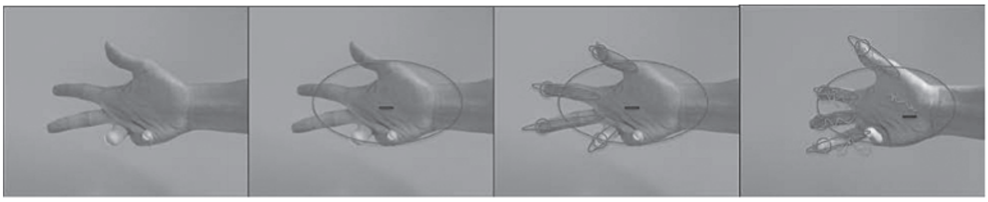


Figure 1.3 Steps of hand and fingers motion estimation

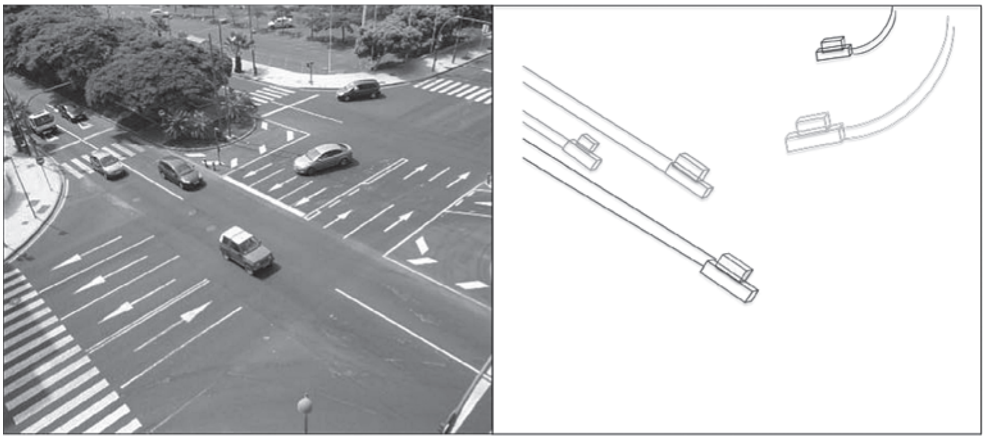


Figure 1.4 The output of a tracking system

1.5.3 Security Applications

Registration methods find many applications in systems used to increase the security for persons and vehicles. Video processing and motion estimation can be used to protect humans from both active and passive accidents. The tracking of vehicles has many potential applications, including road traffic monitoring (see Figure 1.4), digital rear-view mirror, monitoring of car parks and other high-risk or high-security sites. The benefits are equally wide ranging. Tracking cars in roads could make it easier to detect accidents, potentially cutting down the time it takes for the emergency services to arrive.

Another application of motion interpretation in image sequences is a driver assistance system for vehicles driving on the highway. The aim is to develop a digital rear-view mirror. This device could inform the driver when a lane-shift is unsafe. A single camera-based sensor can be used to retrieve information of the vehicle's environment in question. Vehicles driving behind the vehicle in question limit its motion possibilities [14]. Therefore, the application needs to estimate their motion relative to the first car. This problem is illustrated in Figure 1.5.

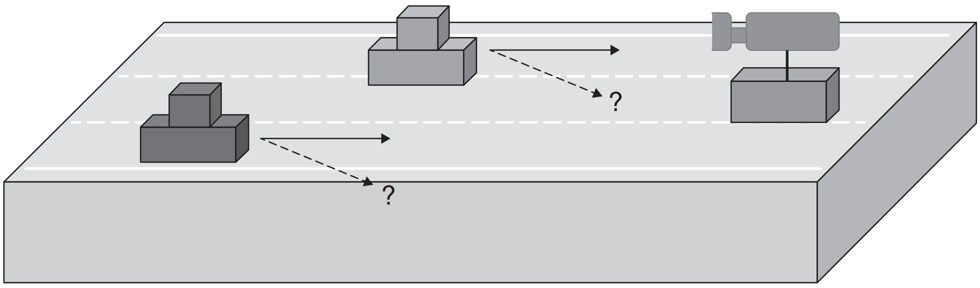


Figure 1.5 Observing other vehicles with a single camera

Figure 1.6 (a) Intelligent CCTV system, (b) 3D indoor visual system

Tracking human motion can be useful for security and demographic applications. For example, an intelligent CCTV system (see Figure 1.6(a)) will be able to ‘see’ and ‘understand’ the environment. In shopping centres, it will be able to count the customers and to provide useful information (e.g. number of customers on a particular day or in the last 2 h). Robots can be utilised for indoor applications without the need of prior knowledge of the building being able to move and perform specific operations. This kind of application can be used in hospitals or offices where the passages can be easily identified (see Figure 1.6(b)).

Visual tracking is an interesting area of computer vision with many practical applications. There are good reasons to track a wide variety of objects, including aeroplanes, missiles, vehicles, people, animals and microorganisms. While tracking single objects alone in images has received considerable attention, tracking multiple objects simultaneously is both more useful and more problematic. It is more useful since the objects to be tracked often exist in close proximity to other similar objects. It is more problematic since the objects of interest can touch, occlude and interact with each other; they can also enter and leave the scene.

The above-mentioned applications are based on machine vision technology, which utilises an imaging system and a computer to analyse the sequences and take decisions. There are two basic types of machine vision applications – inspection and control. In inspection applications, the machine vision optics and the imaging system enable the processor to 'see' objects precisely and thus make valid decisions about which parts pass and which parts must be scrapped. In control applications, sophisticated optics and software are used to direct the manufacturing process.

As it was shown in these examples, object tracking is an important task within the field of computer vision. In Chapter 3, the concept of optical flow for tracking and activity recognition are analysed presenting the related registration methodologies.

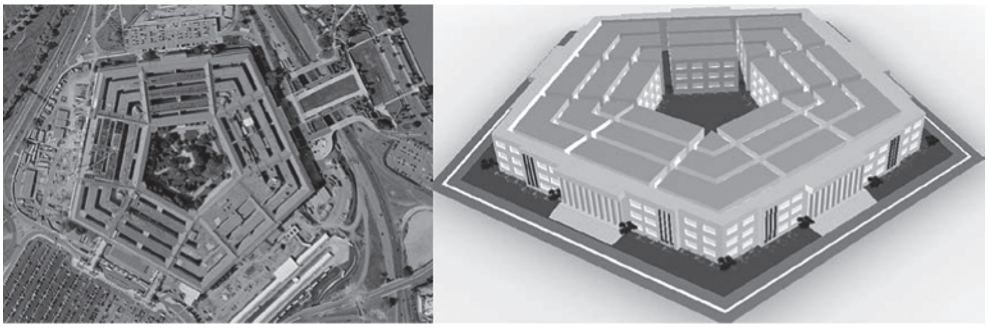
Considering a security system, another important application is face tracking and recognition. It is a biometric system for automatically identifying and verifying a person from a video sequence or frame. It is now very common to find security cameras in airports, offices, banks, ATMs and universities and in any place with an installed security system. Face recognition should be able to detect initially a face in an image. Then features are extracted that are used to recognize the human, taking into account factors such as lighting, expression, ageing, illumination, transformation and pose. Registration methods are used especially for face alignment tasks and they are presented in Chapter 4 highlighting the main problems, the different approaches and metrics for the evaluation of these tasks.

1.5.4 Military and Satellite Applications

Military applications are probably one of the largest areas for computer vision, even though only a small part of the work is open to the public. The obvious examples are detection of enemy soldiers or vehicles and guidance of missiles to a designated target. More advanced systems for missile guidance send the missile to an area rather than a specific target, and target selection is made when the missile reaches the area based on locally acquired image data. Modern military concepts, such as 'battlefield awareness', imply that various sensors, including image sensors, provide a rich set of information about a combat scene, which can be used to support strategic decisions. In this case, automatic data processing is used to reduce complexity and to fuse information from multiple sensors increasing reliability.

Night and heat vision can be used by police to hunt a criminal. Motion estimation techniques and object tracking systems use these vision systems to obtain better performance at night and in cold areas. Torpedoes, bombs and missiles use motion estimation to find and follow a target. Also, motion estimation is utilised by aeroplanes, ships and submarines along with object tracking systems.

Satellite object tracking is one of the military applications with the most research programmes. In this case, the systems track objects such as vehicles, trains, aeroplanes or ships. The main problem in these applications is the low quality of the pictures. Therefore, these practical problems can be reduced when efficient algorithms are used.



(a) (b)

Figure 1.7 (a). Fragment of the satellite image, (b) 3D visualization of the part of the city, obtained as the result of the high-resolution satellite image processing. *Source:* http://commons.wikimedia.org/wiki/File:Pentagon-USGS-highres-cc.jpg

Satellite high-resolution images can also be used to produce maps and 3D visualization of territory status of a city (see Figure 1.7). In this case, image alignment is essential to obtain accurate maps and visualizations; therefore, sub-pixel registration methods are required. Also, due to the size of the captured data, fast approaches are crucial for this type of military and satellite applications. In Chapter 5, a detailed analysis of satellite image registration is presented. Methods for evaluating the performance of these techniques in the context of achieving specific performance goals are also described.

1.5.5 Reconstruction Applications

A fundamental task in computer vision is the one referred to as image-based 3D reconstruction for the creation of 3D models of a real scene from 2D images. Three-dimensional digital models are used in applications such as animation, visualization and navigation. One of the most common techniques for the image-based 3D reconstruction is structure from motion due to its conceptual simplicity. Structure from motion is a mechanism of simultaneously estimating the 3D geometry of a scene (structure) and the camera location (3D motion). Structure from motion is applied, for example, to reconstruct 3D archaeological building and statues using 2D images (see Figure 1.8).

In these approaches, the first step is to find the correspondence of sparse features among consecutive images using feature extraction and matching techniques. In the second step, structure from motion is applied in order to obtain the 3D shape and the motion from the camera. In the final step, both reconstruction and motion are adjusted and refined. Structure from motion is also used in different scenarios, such us autonomous navigation and guidance, augmented reality, hand/eyes motion capture, calibration, remote sensing and segmentation.

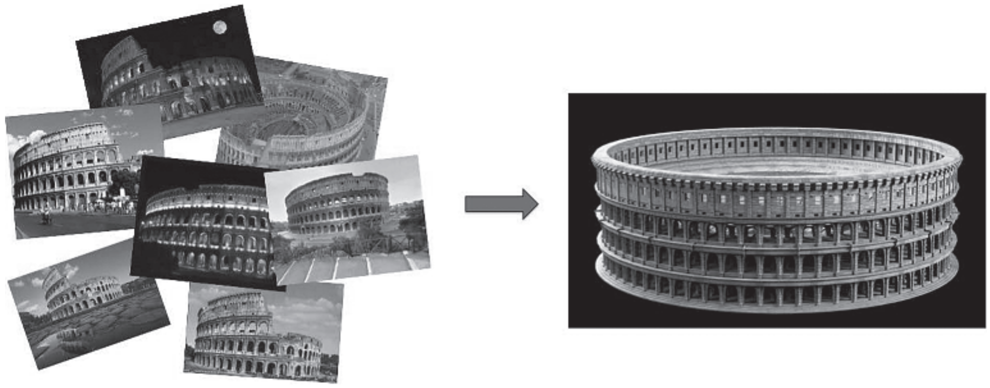


Figure 1.8 The 3D shape of a monument reconstructed using a sequence of 2D images

In Chapter 6, methods for obtaining the 3D shape of an object or an area using motion information and registration techniques are presented. Also registration methods for panoramic view for digital cameras and robotics are described focusing also on issues related to performance and computational complexity.

1.6 Organization of the Book

This book will provide an analysis on registration methodologies and quality metrics covering the most important research areas and applications. It is organized as follows:

Chapter 1 An introduction to the concepts of image and video registration is presented including examples of the related applications. An historical overview on image registration and the fundamentals on quality assessment metrics are also analysed.

Chapter 2 An overview of block-matching motion estimation methods is presented including traditional methods such as full search, tree steps, diamond and other state-of-the-art approaches. The same structure is used for hierarchical and shape-adaptive methods. The concepts of quality of system (QoS) and quality of experience (QoE) are discussed, and how image and video quality is influenced by the registration techniques is analysed. Quality metrics are presented focusing on coding applications.

Chapter 3 The concept of optical flow for tracking and activity recognition is analysed presenting the related state-of-the-art methods indicating their advantages and disadvantages. Approaches to extract the background and track objects and humans for surveillance are presented. Also, methods that operate in real time and focus on activity recognition both for security and entertainment (TV games) are discussed. The concepts of multitarget tracking evaluation including methods for detection recall, detection precision, number of ground truth detections, number of correct detections, number of false detections, number of total detections, and number of total tracks in ground truth are presented. Regarding action recognition metrics related to performance, time accuracy and quality of execution are also analysed.

Chapter 4 Registration methods for face alignment utilised mainly for recognition are presented. Both 2D and 3D techniques are included and different features are discussed. Problems that may affect the registration problem are overviewed including illumination and pose variations, occlusions and so on. Also, metrics for evaluation of these parts of the overall face recognition process are presented and analysed separately and as a whole.

Chapter 5 Methods for global motion estimation are introduced in this chapter operating both in the pixel and Fourier domains. The registration problem is further extended to rotation and scale estimation for satellite imaging applications. Methods for evaluating the performance of these techniques in the context of achieving specific performance are presented. Also the observed quality of registration methods is discussed.

Chapter 6 Methods for obtaining the 3D shape of an object or a scene using motion information and registration techniques are analysed. Additionally, registration methods for panoramic view for digital cameras and robotics are presented focusing on issues related to performance and computational complexity. Also the issues of performance and quality are addressed taking into consideration the related applications.

Chapter 7 The main problems of medical image data analysis are discussed. The particular characteristics of this type of data and the special requirements based on the exact application affect the registration methods. A variety of techniques are discussed applied on either 2D or 3D data. Also intra- and inter-subject image registration methods are reviewed such as non-rigid brain warping. Regarding the quality part, a similar structure with the other chapters is considered.

Chapter 8 Methods for video stabilization in real time based on motion estimation are presented. Furthermore, approaches to remove dirt, flickering and other effects during the restoration process of a video sequence based on registration are analysed, focusing on the latest approaches and the state-of-the-art techniques. Mechanisms for evaluating the restoration accuracy of algorithms operating in either real time or offline are discussed focusing on the quality of experience.

References

[1] Roberts, L.G. (1963) Machine perception of 3-D solids. Ph.D. Thesis. MIT.

[2] Barnea, D.I. and Silverman, H.F. (1972) A class of algorithms for fast digital image registration. *IEEE Transactions on Computers*, **21** (2), 179–186.

[3] Anuta, P.E. (1969) Registration of multispectral video imagery. *Society of Photographic Instrumentation Engineers Journal*, **7**, 168–175.

[4] Anuta, P.E. (1970) Spatial registration of multispectral and multitemporal digital imagery using fast Fourier transform techniques. *IEEE Transactions on Geoscience Electronics*, **8** (4), 353–368.

[5] Leese, J.A., Novak, G.S. and Clark, B.B. (1971) An automatic technique for obtaining cloud motion from geosynchronous satellite data using cross correlation. *Applied Meteorology*, **10**, 110–132.

[6] Pratt, W.K. (1974) Correlation techniques for image registration. *IEEE Transactions on Aerospace and Electronic Systems*, **10** (3), 353–358.

[7] Fischler, M.A. and Elschlager, R.A. (1973) The representation and matching of pictorial structures. *IEEE Transactions on Computers*, **22**, 67–92.

[8] Bajcsy, R., Lieberson, R. and Reivich, M. (1983) A computerized system for the elastic matching of deformed radiographic images to idealized atlas images. *Journal of Computer Assisted Tomography*, **7** (4), 618–625.

[9] Bohm, C., Greitz, T., Berggren, B.M. and Olsson, L. (1983) Adjustable computerized stereotaxic brain atlas for transmission and emission tomography. *American Journal of Neuroradiology*, **4**, 731–733.

[10] Zitova, B. and Flusser, J. (2003) Image registration methods: a survey. *Image and Video Computing*, **21**, 977–1000.

[11] Wang, Z., Bovik, A., Sheikh, H. and Simoncelli, E. (2004) Image quality assessment: from error measurement to structural similarity. *IEEE Transactions on Image Processing*, **13** (4), 600–612.

[12] Sheikh, R., Sabir, M. and Bovik, A.C. (2006) A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, **15** (11), 3440–3451.

[13] Bretzner, L. and Lindeberg, T. (2000) Qualitative multi-scale feature hierarchies for object tracking. *Journal of Visual Communication and Image Representation*, **11**, 115–129.

[14] Leeuwen, M.B. and Groen, A. (2005) Motion estimation and interpretation with mobile vision systems. http://www.science.uva.nl/research/ias/research/perception/vehicmotion/.