R

Radiance

Fabian Langguth and Michael Goesele
GCC - Graphics, Capture and Massively Parallel Computing, TU Darmstadt, Darmstadt, Germany

Related Concepts

►Irradiance

Definition

Radiance L is defined as the power received or emitted per unit solid angle $d\omega$ and per unit projected area $dA\cos\theta$ at a point $\boldsymbol{x}$ in direction (θ, ϕ):

$$L(\boldsymbol{x}, \theta, \phi) = \frac{d^2\Phi}{d\omega dA\cos\theta}.$$

Φ is the radiant power, which describes the total amount of energy that flows through a surface per time interval [1]. θ is the angle between the surface normal of the differential area dA and the direction under consideration (see Fig. 1). The unit of radiance is watt per steradian per square meter ($\mathrm{W}\cdot\mathrm{sr}^{-1}\cdot\mathrm{m}^{-2}$).

Background

Radiance is a concept from radiometry, the science of measuring radiant energy transfer. It is an essential concept for global illumination algorithms [1]. In computer vision, it is mostly used in the context of light transport in scenes [2]. The equivalent concept in photometry is luminance, with the key difference being that luminance is adjusted to account for the varying sensitivity of the human eye to different wavelengths of light.

Theory

An important aspect of radiance is that it is constant along a line of sight in an empty medium. This follows from the conservation of energy in a small bundle of light rays between two differential surface patches. More formally, given two points $\boldsymbol{x}$ and $\boldsymbol{y}$ at distance $r = |\boldsymbol{x} - \boldsymbol{y}|$ and two small areas dx and dy located at $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively, the radiance leaving dx towards dy is written as

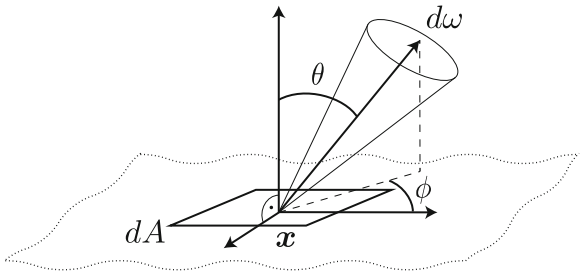
$$L(\boldsymbol{x} \to \boldsymbol{y}) = \frac{d^2\Phi}{d\omega_{\boldsymbol{x}\leftarrow dy}\cos\theta_{\boldsymbol{x}} dx}. \tag{1}$$

$d\omega_{\boldsymbol{x}\leftarrow dy}$ denotes the solid angle covered by dy as seen from $\boldsymbol{x}$. Similarly the radiance arriving at dy from dx is written as

$$L(\boldsymbol{y} \leftarrow \boldsymbol{x}) = \frac{d^2\Phi}{d\omega_{\boldsymbol{y}\leftarrow dx}\cos\theta_{\boldsymbol{y}} dy}. \tag{2}$$

Considering a total vacuum and no additional source of power, all energy in the system must be conserved. This means that all energy leaving from dx towards dy must arrive at dy. Therefore $d^2\Phi$ is equal in both equations, and we can derive

K. Ikeuchi (ed.), *Computer Vision*, DOI 10.1007/978-0-387-31439-6,



Radiance, Fig. 1 Geometric setting

$$L(\boldsymbol{x} \to \boldsymbol{y}) = \frac{d^2\Phi}{d\omega_{\boldsymbol{x}\leftarrow dy} \cos\theta_{\boldsymbol{x}} dx} \tag{3}$$

$$= \frac{d^2\Phi}{\frac{\cos\theta_{\boldsymbol{y}} dy}{r^2} \cos\theta_{\boldsymbol{x}} dx} = \frac{d^2\Phi}{\frac{\cos\theta_{\boldsymbol{x}} dx}{r^2} \cos\theta_{\boldsymbol{y}} dy} \tag{4}$$

$$= \frac{d^2\Phi}{d\omega_{\boldsymbol{y}\leftarrow dx} \cos\theta_{\boldsymbol{y}} dy} = L(\boldsymbol{y} \leftarrow \boldsymbol{x}). \tag{5}$$

Application

An application of the fact that radiance is constant along a ray can be seen in photography. The correct exposure is hereby determined by measuring the radiance (or more precisely the luminance) of a scene or part of a scene that should be correctly exposed. Once exposure is determined, the scene can be photographed from an arbitrary distance without changing the exposure settings of the camera.

References

1. Dutré P, Bala K, Bekaert P (2006) Advanced global illumination. AK Peters, Wellesley
2. Seitz SM, Matsushita Y, Kutulakos KN (2005) A theory of inverse light transport. In: Proceedings of the 10th IEEE international conference on computer vision (ICCV 2005), Beijing

Radiometric Calibration

Yasuyuki Matsushita
Microsoft Research, Beijing, China

Synonyms

Radiometric camera calibration

Related Concepts

▸Radiance

Definition

Radiometric calibration is a process of determining radiometric response functions, which relate sensor irradiance with measured intensity values.

Background

Many computer vision algorithms rely on the assumption that image intensities are linearly related to the image irradiance recorded at the camera sensor. Since most cameras nonlinearly alter irradiance values for purposes such as dynamic range compression, this assumption generally does not hold. It is therefore important to calibrate the response function of a camera, so that the nonlinear mapping can be inverted and subsequent algorithms can assume linearity of intensity observations.

Radiometric calibration aims to estimate the response function f of a camera. The radiometric response function f maps the irradiance I that is captured at the sensor to the image intensity M that is read from the camera:

$$M = f(I). \tag{1}$$

For vision algorithms that require irradiance values I rather than measured intensity M as input, the inverse response function $g = f^{-1}$ needs to be determined so that measured intensities can be made linear with respect to irradiances. Since response functions f are monotonic and continuous, they are invertible to uniquely determine inverse response functions g.

Theory

Radiometric calibration methods require means to collect samples of the radiometric response function with some known relationship. One traditional approach is to use a special target, such as a Macbeth color chart [1], which has color patches with known reflectances. By uniformly illuminating the target, known radiances from the color patches are recorded.

The radiometric response function is then obtained by relating the sensor irradiance I with the recorded intensity values M. Nayar and Mitsunaga [13] use an optical filter with spatially varying transmittance; the variation corresponds to the radiance ratio.

To avoid using such special equipment, some methods use a set of images of a static scene from a fixed viewpoint, taken with different exposure times, so the radiance ratio is known. Known exposure times provide information about sensor irradiance ratios, which are the ratios of exposure times. In a similiar manner with the approach of using a special target, by relating the sensor irradiance I with the measured intensities M, a radiometric response function is estimated. The early work of Mann and Picard [10] uses a gamma correcting function to represent response functions. With known exposure ratios, their method can successfully recover the inverse response function in the parametric form. With only approximate knowledge of relative exposure levels, Mitsunaga and Nayar [12] iteratively solve for a response function based on the assumption that it has a polynomial form. Other iterative estimation methods include that of Tsin et al [18], which estimates nonparametric responses using a statistical model of the CCD imaging process, and that of Pal et al [14], which utilizes probabilistic imaging models and prior models of response functions to compute response functions that can differ from image to image. Debevec and Malik [2] assumed a smoothness property of the response functions and estimated them in a nonparametric manner. As pointed out in [3, 5, 8], without the knowledge of exposure ratios, the estimate still has an *exponential ambiguity*. While not unique, such an estimate is still useful for many applications, such as radiometric alignment, high-dynamic range image production, and image stitching.

Several methods have been developed that use multiple exposures but do not require precise registration. Grossberg and Nayar [3] use the relationship between the intensity histograms of two scenes imaged with different exposures, because intensity histograms are relatively unaffected by small changes in the scene. Kim and Pollefeys [5] compute point correspondences between images. Mann [9] estimates response functions from a rotating and zooming camera.

Instead of using varying exposure times, some approaches use statistical or physical properties embedded in images to achieve radiometric calibration. Tsin et al's method [18] estimates nonparametric response functions using a statistical model of the CCD imaging process. Pal et al [14] used probabilistic imaging models and weak prior models for deriving response functions to produce high-quality high dynamic range images. Matsushita and Lin [11] proposed to use the symmetric property of image noise by observing noise distributions contained in images. Takamatsu et al [16, 17] improved the noise-based method with a probabilistic intensity similarity measure, which requires a fewer number of images. Lin et al [6] and Lin and Zhang [7] proposed a method that takes only a single image as input. Their method uses edges for obtaining color or gray-scale histogram distributions, and the optimal inverse response function is determined by transforming linear distributions. Their method uses a database of response functions (DoRF) compiled by Grossberg and Nayar [4]. In a similar manner, Wilburn et al [19] use temporal color mixtures to directly sample the response function by observing motion blur in an image. More recently, Shi et al [15] show a calibration method from images taken under varying lighting conditions. In their approach, an inverse response function is determined by linearizing color profiles that are defined as a set of measured RGB values at a pixel across images.

Open Problem

In many practical situations, the input dataset is naturally restricted by the camera's capability or application scenarios. For example, it is difficult to obtain multiple images at different exposures with ordinary web cameras. A more general and robust approach for radiometric calibration is still to be investigated.

References

1. Chang Y-C, Reid JF (1996) Rgb calibration for color image analysis in machine vision. IEEE Trans Image Process 5(10):1414–1422
2. Debevec P, Malik J (1997) Recovering high dynamic range radiance maps from photographs. In: Proceedings of ACM SIGGRAPH. Association for Computing Machinery, New York, pp 369–378
3. Grossberg M, Nayar S (2003) Determining the camera response from images: what is knowable? IEEE Trans Pattern Anal Mach Intell 25(11):1455–1467
4. Grossberg M, Nayar S (2003) What is the space of camera response functions? In: Proceedings of computer vision and pattern recognition (CVPR). IEEE Computer Society, Los Alamitos, pp 602–609
5. Kim S-J, Pollefeys M (2004) Radiometric alignment of image sequences. In: Proceedings of computer vision and

pattern recognition (CVPR). IEEE Computer Society, Los Alamitos, pp 645–651
6. Lin S, Gu J, Yamazaki S, Shum H-Y (2004) Radiometric calibration from a single image. In: Proceedings of computer vision and pattern recognition (CVPR). IEEE Computer Society, Los Alamitos, pp 938–945
7. Lin S, Zhang L (2005) Determining the radiometric response function from a single grayscale image. In: Proceedings of computer vision and pattern recognition (CVPR). IEEE Computer Society, Los Alamitos, pp 66–73
8. Litvinov A, Schechner YY (2005) Addressing radiometric nonidealities: a unified framework. In: Proceedings of computer vision and pattern recognition (CVPR). IEEE Computer Society, Los Alamitos, pp 52–59
9. Mann S (2001) Comparametric imaging: estimating both the unknown response and the unknown set of exposures in a plurality of differently exposed images. In: Proceedings of computer vision and pattern recognition (CVPR). IEEE Computer Society, Los Alamitos, pp 842–849
10. Mann S, Picard RW (1995) On being 'undigital' with digital cameras: extending dynamic range by combining differently exposed pictures. In: Proceedings of the IS & T, 48th annual conference. The Society for Imaging Science and Technology, Springfield, pp 422–428
11. Matsushita Y, Lin S (2007) Radiometric calibration from noise distributions. In: Proceedings of computer vision and pattern recognition (CVPR). IEEE Computer Society, Piscataway, pp 1–8
12. Mitsunaga T, Nayar S (1999) Radiometric self-calibration. In: Proceedings of computer vision and pattern recognition (CVPR). IEEE Computer Society, Los Alamitos, pp 374–380
13. Nayar S, Mitsunaga T (2000) High dynamic range imaging: spatially varying pixel exposures. In: Proceedings of computer vision and pattern recognition (CVPR). IEEE Computer Society, Los Alamitos, pp 472–479
14. Pal C, Szeliski R, Uyttendale M, Jojic N (2004) Probability models for high dynamic range imaging. In: Proceedings of computer vision and pattern recognition (CVPR). IEEE Computer Society, Los Alamitos, pp 173–180
15. Shi B, Matsushita Y, Wei Y, Xu C, Tan P (2010) Self-calibrating photometric stereo. In: Proceedings of computer vision and pattern recognition (CVPR). IEEE, New York, pp 1118–1125
16. Takamatsu J, Matsushita Y, Ikeuchi K (2008) Estimating camera response functions using probabilistic intensity similarity. In: Proceedings of computer vision and pattern recognition (CVPR). IEEE Computer Society, Piscataway, pp 1–8
17. Takamatsu J, Matsushita Y, Ikeuchi K (2008) Estimating radiometric response functions from image noise variance. In: Proceedings of European conference on computer vision (ECCV). Springer, Berlin, pp 623–637
18. Tsin Y, Ramesh V, Kanade T (2001) Statistical calibration of ccd imaging process. In: Proceedings of international conference on computer vision (ICCV). IEEE Computer Society, Los Alamitos, pp 480–487
19. Wilburn B, Xu H, Matsushita Y (2008) Radiometric calibration using temporal irradiance mixtures. In: Proceedings of computer vision and pattern recognition (CVPR). IEEE Computer Society, Piscataway

Radiometric Camera Calibration

▸Radiometric Calibration

Radiometric Response Function

Yasuyuki Matsushita
Microsoft Research, Beijing, China

Synonyms

Camera response function

Related Concepts

▸Radiance; ▸Radiometric Calibration; ▸Vignetting

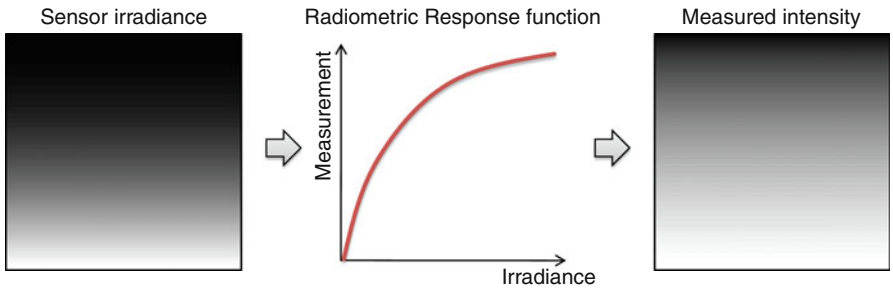
Definition

Radiometric response function is a function that transforms sensor irradiance into measured intensities that are the output from the camera.

Background

In most cameras, there exists a radiometric response function that relates sensor irradiance to measured intensity values. The radiometric response functions are typically nonlinear. This nonlinearity is intentionally designed by camera manufacturers for purposes such as compressing the dynamic range of scene brightness or to take into account the nonlinear mapping of display systems.

While many computer vision algorithms assume a linear (or affine) relationship between the sensor irradiance and the measured image intensity, the radiometric response functions are typically unknown, and these vary with camera parameter settings. Therefore, it is important to estimate the response function to linearize the measured image intensity values for vision algorithms to work. The process of determining radiometric response functions is called *radiometric*



Radiometric Response Function, Fig. 1 A radiometric response function relates the incoming sensor irradiance to measured intensity values

calibration. Once the radiometric response function is determined, the measured intensity can be linearized and transformed into sensor irradiance with a scaling ambiguity.

Theory

The radiometric response function f maps irradiance I that is captured at the sensor to the image intensity M by

$$M = f(I). \tag{1}$$

For computer vision algorithms that require irradiance values I rather than measured intensity M as input, the inverse response function $g = f^{-1}$ needs to be determined so that measured intensities can be made linear with respect to irradiances. Since response functions f are continuous and monotonic, they are invertible to determine inverse response functions (Fig. 1).

Representation

Many parametrization methods have been used to represent radiometric response functions f. To deal with the scale difference between irradiance I and measured intensity M, both are normalized in the range of $[0, 1]$ so that $f(0) = 0$ and $f(1) = 1$.

When representing a radiometric response function in a certain form, there is a trade-off between complexity and flexibility. A simpler representation makes the estimation problem more tractable at the cost of approximation accuracy. On the other hand, a more flexible representation requires an uneasy solution method. Major representations of radiometric response functions are reviewed here.

Mann and Picard [6] represent the response functions in a form of a gamma correction function as

$$M = f(I) = \alpha + \beta I^{\gamma}, \tag{2}$$

where α and β are offset and scale factors, and γ is the power-law parameter.

Mitsunaga and Nayar use a high-order (order of N) polynomial function as the model of inverse response functions g as

$$I = g(M) = \sum_{n=0}^{N} c_n M^n, \tag{3}$$

where c_n are the coefficients of the polynomial function.

Grossberg and Nayar apply principal component analysis (PCA) to a database of real-world response functions (DoRF) and show that the space of response functions can be represented by a small number of basis functions [3].

$$g = \bar{g} + \sum_{n=1}^{N} c_n g_n. \tag{4}$$

In the above equation, $\bar{g}$ is the mean inverse response, and g_i is the i-th principal component of the inverse response functions.

Debevec and Malik [1] use a nonparametric form of the radiometric response functions. The nonparametric representation has great descriptive power, but it is necessary to estimate $f(I)$ (or $g(M)$) for each intensity level (e.g, 256 for 8-bit images). Therefore, the solution methods tend to become more complex. The same representation is also used by Tsin et al [11].

Existing methods can estimate radiometric response functions from a set of images taken with different known exposure times from a fixed viewpoint [1, 2, 8]. More recent methods use different cues to achieve the estimation in more general settings. These include methods using single-image edges [4, 5], image noise observations [7, 9, 10], and motion blur [12]. These estimation methods are detailed in the entry "▸Radiometric Calibration".

Application

The estimation of radiometric response functions constitutes an extensively researched area due to its fundamental importance for many computer vision algorithms, such as shape-from-shading, photometric stereo, high-dynamic-range imaging, and photo stitching. Estimated radiometric response functions are used to linearize the measured intensity values as preprocessing for these computer vision algorithms that require a linear (or affine) relationship with the irradiance.

References

1. Debevec PE, Malik J (1997) Recovering high dynamic range radiance maps from photographs. In: Proceedings of ACM SIGGRAPH. ACM SIGGRAPH, New York, pp 369–378
2. Grossberg MD, Nayar SK (2003) Determining the camera response from images: what is knowable? IEEE Trans Pattern Anal Mach Intell (PAMI) 25(11):1455–1467
3. Grossberg MD, Nayar SK (2004) Modeling the space of camera response functions. IEEE Trans Pattern Anal Mach Intell (PAMI) 26(10):1272–1282
4. Lin S, Gu J, Yamazaki S, Shum H-Y (2004) Radiometric calibration from a single image. In: Proceedings of computer vision and pattern recognition (CVPR), vol 2. IEEE Computer Society, Los Alamitos, pp 938–945
5. Lin S, Zhang L (2005) Determining the radiometric response function from a single grayscale image. In: Proceedings of computer vision and pattern recognition (CVPR), vol 2. IEEE Computer Society, Los Alamitos, pp 66–73
6. Mann S, Picard RW (1995) Being eundigitalf with digital cameras: Extending dynamic range by combining differently exposed pictures. In: Proceedings of IS & T 48th annual conference. The Society for Imaging Science and Technology, Springfield, pp 422–428
7. Matsushita Y, Lin S (2007) Radiometric calibration from noise distributions. In: Proceedings of computer vision and pattern recognition (CVPR). IEEE Computer Society, Los Alamitos
8. Mitsunaga T, Nayar SK (1999) Radiometric self-calibration. In: Proceedings of computer vision and pattern recognition (CVPR), vol 2. IEEE Computer Society, Los Alamitos, pp 374–380
9. Takamatsu J, Matsushita Y, Ikeuchi K (2008) Estimating camera response functions using probabilistic intensity similarity. In: Proceedings of computer vision and pattern recognition (CVPR). IEEE Computer Society, Los Alamitos
10. Takamatsu J, Matsushita Y, Ikeuchi K (2008) Estimating radiometric response functions from image noise variance. In: Proceedings of European conference on computer vision (ECCV). Springer, New York, pp 623–637
11. Tsin Y, Ramesh V, Kanade T (2001) Statistical calibration of ccd imaging process. In: Proceedings of international conference on computer vision (ICCV), vol 1. IEEE Computer Society, Los Alamitos, pp 480–487
12. Wilburn B, Xu H, Matsushita Y (2008) Radiometric calibration using temporal irradiance mixtures. In: Proceedings of computer vision and pattern recognition (CVPR). IEEE Computer Society, Los Alamitos

Rao Metric

▸Fisher-Rao Metric

Rationale for Computational Vision

Ruzena Bajcsy
Department of Electrical and Computer Sciences, College of Engineering University of California, Berkeley, CA, USA

Synonyms

Computer vision

Definition

Vision is a scientific field that investigates biological systems and machines how to use light to gain information about their environments. It covers several subfields such as optics, perception, psychophysics, neurophysiology, information science, signal processing, cognitive science, and related subjects.

Background

Background section will not be a review of all the contributions to the field of vision from all the subfields mentioned above. Rather it shall concentrate on how different subfields try to solve the problem vision and

how this field has evolved over time due to better understanding of the problems but also due to more powerful technological tools.

Fundamentally vision utilizes the spatial and temporal information (structure) that stems from the reflection of light from the environment. The focus will be on the computational aspects of processing of this visual information, asking how visual information is represented for recognition, mobility, and manipulation, hence computational vision.

Much of the machine vision has been motivated by various applications: military medical, industrial, cultural, and commercial. The application-driven solution can be a separate article by itself. In this article the emphasis will be more on the scientific rational for computational vision than all the other reasons.

Motivation for this scientific endeavor comes from two very different sources:

1. How to model (mathematically and/or algorithmically) the biological process of vision. Can one design vision-based processes that explain visual perceptual phenomena?
2. How to design machines that will produce the desired outcome (recognition or navigation of a robot) from visual data.

The goals and the evaluation of success of these two approaches are very different.

Computational models of biological systems and machine vision is a very broad subject because it entails mammals as well as humans. It also encompasses fields such as computational neuroscience, through psychophysics up to computational cognitive modeling. In this article only some representative works will be mentioned which in no way does justice to this field. It covers modeling at different scale from neurons with prominent work of Hodgkin–Huxley model [1] and Hubel and Wiesel [2], through sensory processing and communication of Barlow [3]. Psychologists have seriously considered computational models, especially how to explain visual illusions (exemplars are Gregory [4] and Frisby [5]), mechanisms of binocular vision (see Julesz [6]), and theory of color vision in retinex theory (by Land [7]).

Coincidentally, the engineering community also was heavily engaged in developing computational models for processing data coming from cameras. The limitations of poor resolution of cameras (64 by 64 pixels), limited memory, and compute power limited the applications of this era. Hence the topics were primarily two-dimensional picture recognition, such as digit recognition, finger print recognition, and in medical domain chromosome recognition. The methodologies used were statistical pattern recognition represented by Duda and Hart [8], signal-image processing by Rosenfeld and Kak [9], syntactic pattern recognition by K.S. Fu [10], and many others.

In early 1970s another trend emerged from the newly established Artificial Intelligence Laboratories at MIT and Stanford Universities. This trend emphasized the need for recognizing three-dimensional (3D) objects and scenes, arguing that today beings live in the 3D world; therefore the visual objects need to be represented as such. Guzman [11] was the first one who dealt with this problem using principles of projective geometry. Binford and Agin [12] and his students proposed a generalized cylindrical model (skeleton and corresponding cross-sections) as a generic model of 3D objects. It was argued that the model can represent parts and in turn the part-whole relationships can be represented as graphs.

This was more or less object-centered representation as oppose to the viewer-centered representation, proposed by Koenderink and vanDoorn [13] as aspect graphs. This debate object-centered representation vs. viewer-centered representation is still ongoing!

In meantime David Marr [14] in late 1970s and early 1980s questioned the scientific approach of engineering community as ad hoc and not anchored in scientific theory. He argued that the validation of the engineering approaches must be guided by what is known from neuroscience and psychophysics about human visual processing (this at present is the best model).

He outlined three levels of analysis:

1. The computational theory, specifying the goals of the computation
2. Representation and algorithm, giving a representation of the input and output which transforms one into the other
3. The hardware implementation, how algorithm and representation may be physically realized

Concurrently with Marr's effort in the UK, Christopher Longuet-Higgins [15] introduced cognitive science as the interdisciplinary study of how information is represented and transformed in the brain. It spans many levels of analysis, from low-level learning and decision mechanisms to high-level logic and planning. This field has blossomed with many works of modeling of perceptual and cognitive processes (see Newel and Simon [16], Hinton [17], and many others).

The 1980s benefitted from rapid advancements in hardware both in better cameras and computing power. This development enabled to perform some real-time computation and connect signal processing and perception with action. Motivated by Gibson [18] and Bajcsy [19] proposed a new research paradigm, called Active Perception. In this new framework, it has been shown that there is benefit in controlling the data acquisition and serves as way of modeling focus of attention.

Simultaneously progress has been made in various basic algorithms such as stereo reconstruction, motion detection and interpretation, multiple view reconstruction, shape from shading and photometric stereo, and shape–object–scene representation. Several textbooks cover these advancements, just to cite a few: Faugeras [20], Forsyth and Ponce [21], and Ma et al. [22]:

- Another challenge for machine perception comes from Gestalt psychology founded by Max Wertheimer [23], in the beginning of the twentieth century. The word Gestalt means a unified or meaningful whole. The laws of Gestalt theory of perceptual organization are as follows:
- The Law of Similarity
- The Law of Pragnanz
- The Law of Proximity
- The Law of Continuity
- The Law of Closure

Computationally however, they are hard to define and map into an algorithm since they deal with rather vague concepts such as proximity and similarity. From the beginning of computational vision, this problem has attracted attention, first just building some distance function (Zobrist and Thompson [24]) later looking at these laws as a guide to perceptual organization (Low [25]) and more recently as "image segmentation," the partitioning of an image (or video streams) into sets of pixels that correspond to "objects" or parts of objects. This process is based on bottom-up cues such as similarity of pixel brightness, color, texture, and motion as well as top-down input derived from familiar object categories such as faces. Malik and his coworkers [26] aimed at developing a scientific understanding of grouping, both in the context of human perception and for computer vision, and have shown progress in this area but also its limitations.

At the onset of the twenty-first century, the abundance of images on Internet has spurred an interest in Machine Learning Technologies (MLT) as they apply to object recognition and classification. The latest premier vision conferences ICCV and ECCV have been representative examples of this research. The theoretical development related to this effort is compressed sensing; see Emmanuel Candès and Terence Tao [27]. They discovered important results on the minimum number of data needed to reconstruct an image even though the number data would be deemed insufficient by the Nyquist–Shannon criterion. Further effort in combining the compressed sensing with principal component pursuit is in [28].

In conclusion, the scientific community has seen serious progress in computational vision afforded by technological advances (cameras, computing power) but also by the community mastering much more sophisticated mathematical and computational tools than ever before. The emphasis on robustness, sharing code, and creating standard data bases where different approaches can be tested is very good.

Open problems

The basic representations of visual objects and their dynamics are still open problems. Further, the adaptive nature and flexibility of biological vision is still a dream to be accomplished by the computational vision community

References

1. Hodgkin A, Huxley A (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. J Physiol 117:500–544
2. Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol (Lond) 160:106–154
3. Barlow HB (1961) Possible principles underlying the transformation of sensory messages. In: Rosenblith WA (ed) Sensory communication. MIT, Cambridge, pp 217–234
4. Gregory R (1966, 1997) Eye and brain: the psychology of seeing, 1st edn. Weidenfeld and Nicolson, London (1966); 5th edn. Oxford University Press (1997)
5. Frisby JP (1979) SEEING; illusion, brain and mind. Oxford University Press, Oxford/New York
6. Julesz B (1971) Foundations of cyclopean perception. The University of Chicago Press, Chicago
7. Land EH (1977) The retinex theory of color vision. Sci Am 237:108–128

8. Duda R, Hart P, Stork D (2000) Pattern classification, 2nd edn. Wiley, New York
9. Rosenfeld A, Kak A (1982) Digital picture processing. Academic, New York
10. Fu KS (1981) Syntactic pattern recognition and applications. Prentice Hall, Upper saddle river, New Jersey
11. Guzman A (1968) Decomposition of a visual scene into three-dimensional bodies. Proceedings of AFIPS fall conference in Dec 1968, San Francisco. ACM, New York
12. Agin G, Binford T (1976) Representation and description of curved objects. IEEE Trans Comput 25(4):439–449
13. Koenderink JJ, van Doorn AJ (1979) The internal representation of solid shapes with respect to vision. Biol Cybern 32:211–216
14. Marr D (1982) Vision: a computational investigation into the human representation and processing of visual information. MIT, Cambridge
15. Longuet-Higgins HC (1979) The perception of music. Proc R Soc B 205:307–322
16. Newell A, Simon HA (1972) Human problem solving. Prentice Hall, Englewood Cliffs
17. Hinton GE, Sejnowski TJ (1983) Analyzing cooperative computation. In: Proceedings of the fifth annual conference of the Cognitive Science Society, Rochester
18. Gibson JJ (1987) The ecological approach to visual perception. Lawrence Erlbaum Associates, Hillsdale
19. Bajcsy R (1988) Active perception. Proc IEEE 76(8): 996–1006
20. Faugeras O (1993) Three-dimensional computer vision, a geometric viewpoint. MIT, Cambridge
21. Forsyth DA, Ponce J (2003) Computer vision, a modern approach. Prentice Hall, Upper Saddle River/London
22. Ma Y, Soatto S, Kosecka J, Sastry SS (2004) An invitation to 3-D vision; from images to geometric models. Springer, New York
23. Wertheimer M (2000) Laws of organization of perceptual forms. In: Yantis S (ed) Visual perecption: essential readings. Key readings in cognition. Psychology Press, Philadelphia, pp 216–224
24. Zobrist AL, Thompson WB (1975) Building a distance function for Gestalt grouping. IEEE Trans Comput C-24(7): 718–728
25. Lowe DG (1985) Perceptual organization and visual recognition. Kluwer Academic, Boston
26. Malik J, Arbaleaz P, Maire M (2009) Projects: grouping and ecological statistics. http://www.eecs.berkeley.edu/Research/Projects/CS/vision/vision_group.html
27. Candès EJ, Tao T (2008) Reflections on compressed sensing. IEEE Inf Theory Soc Newsl 58(4):20–23
28. Zhou Z, Wright J, Li X, Candès EJ, Ma Y (2010) Stable principal component pursuit. In: Proceedings of international symposium on information theory, Austin

Recognition-By-Components (RBC) Theory

►Geons

Recovery of Reflectance Properties

Jason Lawrence
Department of Computer Science, School of Engineering and Applied Science University of Virginia, Charlottesville, VA, USA

Synonyms

Appearance scanning; BRDF measurement; Reflectometry

Definition

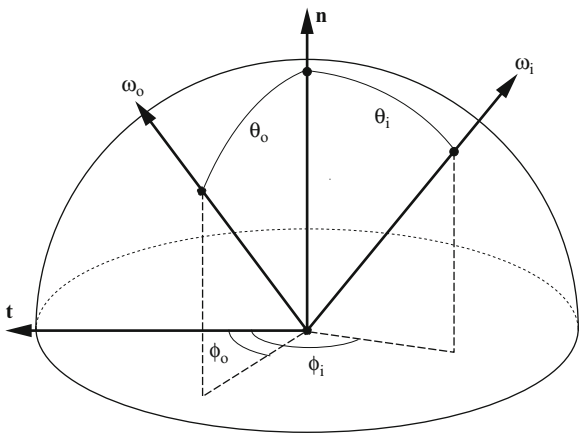
Recovering reflectance properties refers to the process of measuring and modeling the manner in which a material scatters incident light. This often involves estimating the parameters of a light scattering function based on measurements of a physical sample.

Background

Along with its 3D shape, another crucial aspect of an object's appearance is the manner in which it scatters incident light, often referred to generally as its *reflectance*. In the case of opaque objects, it may be assumed that any light which strikes the surface is either reflected back into the environment at that same incident location, possibly in different amounts along different directions, or absorbed by the material. Translucent objects require consideration of the way light scatters internally within the medium. Only metallic surfaces are technically opaque, although dielectrics may be treated as opaque at an appropriate measurement scale. Further, the reflectance of many objects varies spatially over their surface. This entry focuses on techniques for measuring and representing the optical properties of opaque surfaces.

Theory

The appearance of a homogeneous opaque surface is completely characterized by the Bidirectional Reflectance Distribution Function (BRDF) [1]. This is



Recovery of Reflectance Properties, Fig. 1 Geometry of the BRDF. The incident direction ω_i and exitant direction ω_o are defined with respect to the local surface frame defined by the surface normal **n** and tangent **t**. These directions are often written in terms of their respective elevation and azimuthal angles θ and ϕ

a scalar-valued function of the ratio of the differential radiance dI leaving the surface along direction ω_o with respect to the differential surface irradiance dE from light arriving along direction ω_i with wavelength λ

$$f_r(\omega_i, \omega_o, \lambda) = \frac{dI(\omega_o, \lambda)}{dE(\omega_i, \lambda)} \qquad [\text{sr}^{-1}]. \qquad (1)$$

This is a function of five variables (note that the directions are unit length and thus occupy a 2D domain) defined with respect to the local tangent frame attached to a location on a 2-manifold surface (Fig. 1). A very nice overview of the radiometry underlying the BRDF is available in a chapter written by Pat Hanrahan entitled "Rendering Concepts" in the textbook by Cohen and Wallace [2].

Note that Eq. (1) does not differentiate between the wavelength of the incident and exitant light and thus it cannot capture photoluminescence effects such as phosphorescence or fluorescence. Furthermore, the full dependence on wavelength is often ignored (as will be the case in this entry), in which case λ in Eq. (1) is omitted and the BRDF is a function of four variables defined within a trichromatic color space such as RGB. The dimensionality of the BRDF can be further reduced to three if the material's reflectance is *isotropic*, meaning it is unaffected if the incident and exitant directions are rotated together around the surface normal. This is in contrast to *anisotropic* materials which exhibit a visible *surface grain* that causes the reflectance to depend on this azimuthal component such as brushed metal, satin, silk, and velvet.

In many cases, the surface BRDF will vary from one surface location to the next (e.g., a wooden object with visible spatially varying grain densities). The spatially varying BRDF (svBRDF) is used to characterize the reflectance of these inhomogeneous surfaces, and it simply adds surface position **x** to the angular variables in the BRDF: $S(\mathbf{x}, \omega_i, \omega_o)$. The svBRDF is thus a function of six variables since two numbers are required to specify a surface location.

Representations

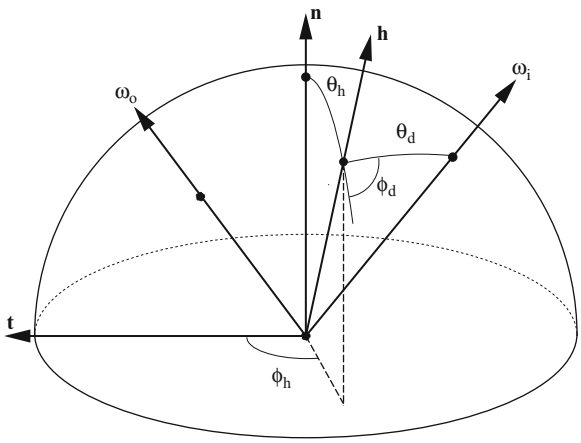
A considerable amount of research has focused on developing efficient BRDF representations that apply to a wide range of materials. This includes a number of early and still widely popular phenomenological models such as the Phong [3] and related Blinn-Phong models [4], which offer reasonable approximations for plastics and smooth painted surfaces.

An alternative approach is to derive the BRDF from the laws of physics based on a hypothesized physical model of a material's surface. These physically based models include the seminal Torrance-Sparrow [5] and Cook-Torrance models [6], both of which assume that the surface is composed of randomly oriented microscopic mirrors, or *microfacets*. These models capture important effects predicted by Fresnel equations including color shifts near highlights and an overall increased specular response near *grazing angles* (i.e., as ω_o and ω_i approach the horizon in opposite directions). The He-Torrance-Sillion-Greenberg (HTSG) model [7] was derived using a similar methodology, but considers wave-related effects such as diffraction and interference. The Oren-Nayar model [8] is also based on a microfacet theory, but one in which the individual microfacets are perfect Lambertian reflectors and are meant to reproduce the reflectance of matte materials such as dust or chalk. The Kajiya-Kay model [9] considers the way light is reflected from small fibers modeled as cylinders and is intended to model the appearance of hair and fur. In all of these cases, the resulting BRDF is an analytic expression with a handful of parameters that control the

magnitude, color, and shape of dominant lobes in the BRDF that are commonly either diffuse (largely nondirectional), specular (forward scattering), or retroreflective (backward scattering).

Another family of BRDFs was derived to fit empirical data. This includes the Lafortune [10] and Ward [11] models. These are both analytic models with parameters similar to those described above. The Lafortune model is defined as the sum of an arbitrary number of cosine lobes and can express a wide range of phenomena. The Ward BRDF is functionally more similar to the Cook-Torrance or Blinn-Phong model, but can produce elongated specular highlights that match a common form of anisotropic reflectance.

More recent research has considered nonparametric representations of the BRDF which are often constructed directly from measured data. These range from straightforward tabulated models [12] to compressed representations obtained by projecting the measurements into either a fixed basis defined over the double hemisphere (e.g., Zernike polynomials, spherical harmonics, wavelets, radial basis functions [13]) or a specialized basis estimated from the measured data itself. In the latter case, this can be achieved using common dimensionality reduction algorithms such as Principal Component Analysis (PCA) or cast as a matrix factorization problem [14, 15]. The accuracy of these representations is greatly affected by the way the BRDF is parameterized. A particularly useful parameterization was introduced by Rusinkiewicz [16] and is based on the *half-angle* and *difference angle* (inset at right). The half-angle is simply the bisector of the incident and exitant directions $\mathbf{h} = (\omega_i + \omega_o)\|\omega_i + \omega_o|$, and the difference angle (θ_d, ϕ_d) is the incident direction expressed with respect to the half-angle. This parameterization has the desirable property of aligning common BRDF features to the transformed coordinate axes – including specular peaks [6], grazing-angle effects, and retroreflective peaks [8] – and as a result, only a relatively small amount of data is needed for an accurate representation.



Acquisition

The most basic requirements of a BRDF measurement system are a light source to illuminate the surface and a photodetector to record the amount of energy reflected in a particular direction. The difficulty of measuring the BRDF of a material lies in the high dimensionality of these functions and the calibration requirements of existing setups.

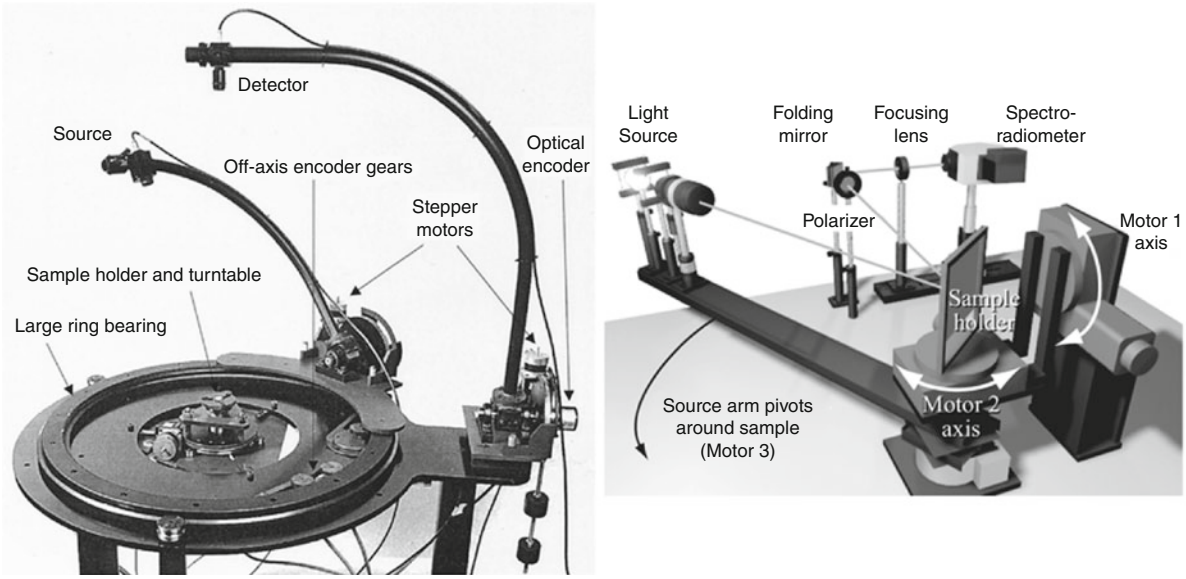
Gonioreflectometers

BRDFs were originally measured using a *gonioreflectometer*, a term that incorporates the Latin word for angle (*gonio*). These consist of a single photodetector and light source that can be moved to arbitrary locations on a hemisphere centered around a small planar sample (Fig. 2). Gonioreflectometers are only suitable for acquiring relatively sparse angular measurements since densely sampling the full BRDF domain would lead to infeasible acquisition times. However, they are very accurate and repeatable and are thus still used to validate newer designs and maintain reflectance standards.

In order to use the data recorded by a gonioreflectometer in a practical system, it is often necessary to estimate the regions of the BRDF that were not directly measured. This is often done by fitting the parameters of one of the analytic BRDF models described previously to the measured data. This process involves solving a nonlinear optimization problem with many variables and is often difficult and error prone in practice [19].

Camera-Based Systems

A seminal development was Greg Ward's use of a curved mirror and camera [11] (Fig. 3). Note that a single image recorded by the camera contains a densely sampled 2D slice of the BRDF at a fixed



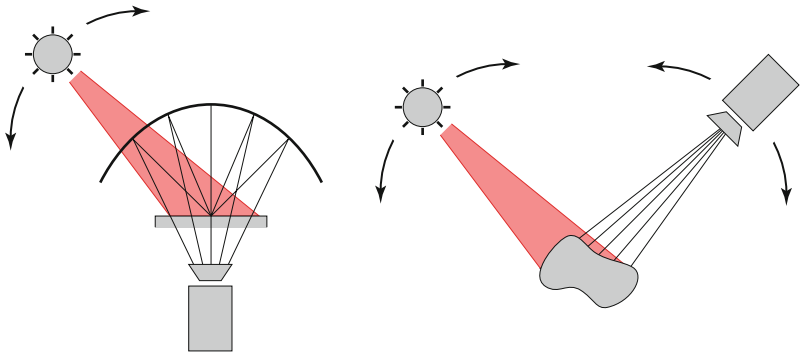
Recovery of Reflectance Properties, Fig. 2 Two gonioreflectometer designs. *Left*: the apparatus developed by White et al. [17] has four degrees of freedom and achieves an angular resolution of approximately 0.1° and is accurate to within 0.3°. (Image reproduced from [17].) *Right*: the design of Li et al. [18] has three degrees of freedom and is thus restricted to measuring isotropic material samples. Due to the use of a spectroradiometer, this system can measure the BRDF at 10 nm increments over the visible spectrum

incident direction. Recording multiple images at different light source positions allows probing the full 4D domain. This setup enables efficiently measuring a considerably larger portion of the BRDF domain than was previously possible with gonioreflectometers and led to the development of a new anisotropic BRDF model that is still in wide use today.

Ghosh et al. [20] describe a related setup that uses a digital projector and multiple curved mirrors to achieve a similar acquisition process, but one that does not require any moving parts, which significantly decreases acquisition time. Dana et al. [21] also explore using a mirrored dome and camera to accelerate BRDF capture. A key aspect of their design is that it allows translating the material sample within the measurement plane in order to sample the spatial dimensions of the svBRDF.

The camera-based BRDF measurement system introduced by Marschner et al. [22] was another significant step in the field, which traded Ward's curved mirror for a curved sample to achieve a similar effect (Fig. 3 right). A unique measurement of the object's BRDF is recorded at each pixel since each pixel observes a different orientation of the surface (recall that the BRDF is defined with respect to the *local* surface frame). Assuming that the sample object's reflectance is homogeneous (i.e., no perceptible spatial variation), this provides a dense 2D slice of its BRDF, although the one that differs from the slices is acquired with Ward's setup. Matusik et al. [12] refined Marschner's design and measured the BRDFs of nearly one hundred isotropic spherical samples at unprecedented angular resolutions. Ngan et al. [19] extended this design further to allow measuring anisotropic samples that were formed into thin strips and wrapped around a cylinder.

Although camera-based systems can achieve a much higher angular resolution than traditional gonioreflectometers, the quality of the individual measurements is generally lower. This is due to the reduced quality of the individual photosensitive elements in a typical CCD or CMOS array, the need for more complex optical systems which can produce internal reflections, and the way wavelength is sampled using, for example, a Bayer filter [23]. However, this is beginning to change with the steady improvement in digital camera technology and the development of tunable narrowband color filters. Additionally, the high angular resolution of the data returned by these systems enables the use of non-parametric representations



Recovery of Reflectance Properties, Fig. 3 Two camera-based BRDF measurement systems. *Left*: Ward's acquisition rig used a curved half-silvered mirror and a camera to enable efficient acquisition of dense BRDF measurements. *Right*: Marschner's setup exploited the same principle through the use of a curved specimen and camera

discussed previously. The advantage of these models is that they often provide a much more accurate fit to the measured data compared to an analytic model and require solving a linear optimization problem which is more stable.

In the aforementioned camera-based systems, the geometry of the target sample and its location with respect to the camera and light source must be known in order to properly interpret the recorded images as BRDF measurements. This calibration step is notoriously difficult and is often simplified by using samples with a specific known geometry: planar, spherical, or cylindrical. A recent research focus has been on developing systems for measuring the BRDF or svBRDF of samples with *arbitrary geometry*. On the one hand, this task is considerably harder since existing techniques for scanning geometry often rely on assumptions about the surface reflectance. On the other hand, all that is needed is the normal direction and, for anisotropic surfaces, the tangent direction at each image pixel as opposed to a complete 3D model.

One of the first general systems for measuring the reflectance of arbitrarily shaped spatially varying opaque isotropic objects was proposed by Lensch et al. [24]. Their approach has multiple steps. First, the geometry of the target object is acquired using either a penetrative method such as a computed tomography (CT) scanner or with a standard laser scanner after first coating the object in a diffuse powder. In a second step, the object is photographed from different viewpoints under variable point lighting (Fig. 3). Third, the 3D geometry in the first step is registered to the images in the second step using a silhouette-based alignment algorithm. The extraction of BRDF data can proceed using the surface normal of the registered 3D model at each image pixel. Finally, they use the resulting BRDF data to estimate the parameters of a Lafortune model [10] at each vertex in a triangle mesh of the surface. To make this fitting process more robust, they assume that the object's reflectance can be accurately modeled with only a handful of unique *basis BRDFs* and corresponding spatial blending weights. Therefore, as opposed to storing a unique set of Lafortune parameters at each vertex, only a set of nonnegative barycentric coordinates (unity of partition) over the set of basis BRDFs is computed instead. Computing these basis BRDFs and blending weights is cast as a clustering problem.

A related system was proposed by Goldman et al. [25]. The key difference is that they *jointly* estimate the surface normal at each pixel in a fixed camera along with the coefficients of the Ward BRDF model by solving a large non-linear optimization problem. On the one hand, this avoids having to scan the geometry in a separate step and subsequently register this geometry to the set of images, significantly simplifying the experimental setup and processing. On the other hand, this approach recovers only the normals and BRDFs for the portion of the object visible from the chosen viewpoint. This approach can be regarded as a generalization of Lambertian photometric stereo [26] since the surface BRDFs are drawn from a much larger space (i.e., the space of BRDFs expressible by the Ward model). Goldman et al. follow a similar strategy as Lensch et al. and assume that object reflectance at each pixel can be expressed as a convex combination

over a small basis of homogeneous BRDFs. They demonstrate the importance of this representation of the svBRDF for achieving a stable optimization.

More recent work has extended the basic approach of Goldman et al. to avoid relying on a parametric BRDF model. The downside to using a parametric model such as the Ward BRDF is that this imposes a particular structure on the surface reflectance. Whenever the object's actual reflectance deviates from what this model can express, errors are introduced. The system presented by Alldrin et al. [27] jointly estimates the surface normal and BRDF at each pixel in a fixed camera where the BRDF is modeled using a tabulated bivariate representation. This can significantly improve the accuracy of the resulting model in many situations.

Another related approach is due to Wang et al. [28] which focuses on measuring the reflectance of anisotropic surfaces. Their system works in two steps. First, they acquire densely sampled angular measurements of the BRDF at a small number of strategically chosen points on the object surface. Second, they record sparse angular measurements sampled densely over the object surface. The sparse measurements in the second step are used to estimate a convex combination of the densely sampled BRDFs in order to recover a model of the svBRDF that is dense in both the spatial and angular dimensions. This basic strategy of combining dense angular data at a small number of surface locations with sparse angular data sampled densely over the surface represents a compelling tradeoff between acquisition time and final model quality. Wang et al. also estimate the tangent direction at each surface location as a by-product of their optimization and model the BRDF using a hybrid representation that combines a tabulated (nonparametric) normal distribution function (NDF) with analytic (parametric) expressions for the shadowing, masking, and Fresnel components [19].

A Note on Bidirectional Texture Functions

This entry has focused on methods for measuring the BRDF or svBRDF of physical objects. There is a family of related techniques that focus on the bidirectional texture function (BTF) [29]. Despite having the same domain as the svBRDF, the BTF conveys a slightly more general notion of reflectance. A BTF quantifies the amount of light that is exchanged (scattered) between pairs of angles located along the surface of a proxy geometry that does not necessarily coincide with the actual surface. As a result, the BTF couples visibility, interreflections, and local reflectance in complex ways that are perceptible at the chosen measurement scale. Technically, the difference between BTFs and svBRDFs is a matter of the degree to which this proxy geometry deviates from the actual object surface, and indeed, no real-world surface is ever perfectly smooth or exhibits exactly the microfacet structure assumed by many BRDFs. Nevertheless, some objects are more suitable than others to be represented as an svBRDF plus a surface, whereas others (e.g., cloth and other hairy or fuzzy surfaces) do not allow resolving the geometry at a fine enough resolution to isolate the local reflectance, and the BTF is perhaps the only option. Systems for measuring BTFs [29, 30] are nearly identical to those for measuring svBRDFs. The difference is a matter of how the resulting data is interpreted.

Application

Measuring and modeling the reflectance of real-world materials is a key component of most graphics and vision systems. For example, recreating a visually rich and realistic virtual 3D world requires populating it with materials with the same variety and intricacy as those found in nature. Similarly, any vision system that attempts to infer information about the 3D structure of a natural image must reason about the way light is absorbed and reradiated by the various surfaces that compose the scene. Therefore, it is important to have efficient and accurate techniques for measuring or *scanning* the reflectance of physical samples.

Open Problems

Developing techniques for measuring and modeling BRDFs and svBRDFs is an active area of research. Much of this work focuses on expanding the set of materials that can be reliably measured to include those with anisotropic and translucent properties. Another thrust of current research focuses on the usability and operation of appearance acquisition systems themselves. Measurement systems will need to be much more efficient and easier to calibrate before they can be reliably deployed in non-laboratory conditions. Finally, extending camera-based systems to measure the spectral dimensions of BRDFs is beginning to

receive serious attention. This includes resolving the spectral profiles of the incident and exitant light in addition to photoluminescence effects caused by materials that absorb energy at one wavelength and emit it at another wavelength.

References

1. Nicodemus FE, Richmond JC, Hsia JJ (1977) Geometrical considerations and reflectance. Natl Bur Stand, NBS MN-160
2. Cohen MF, Wallace JR (1993) Radiosity and realistic image synthesis, 1st edn. Morgan Kaufmann Publishers, San Francisco
3. Phong BT (1975) Illumination for computer generated images. Commun ACM 18:311–317
4. Blinn JF (1978) Simulation of wrinkled surfaces. In: Computer graphics (SIGGRAPH '78 proceedings). ACM, New York, pp 286–292
5. Torrance KE, Sparrow EM (1967) Theory for off-specular reflection from roughened surfaces. J Opt Soc Am 57
6. Cook RL, Torrance KE (1981) A reflectance model for computer graphics. In: Computer graphics (SIGGRAPH 1981). ACM, New York, pp 7–24
7. He XD, Torrance KE, Sillion FX, Greenberg DP (1991) A comprehensive physical model for light reflection. In: Computer graphics (SIGGRAPH 1991). ACM, New York, pp 175–186
8. Oren M, Nayar SK (1994) Generalization of lambert's reflectance model. In: Computer graphics (SIGGRAPH 1994). ACM, New York, pp 239–246
9. Kajiya JT, Kay TL (1989) Rendering fur with three dimensional textures. In: Computer graphics (SIGGRAPH 1989). ACM, New York, pp 271–280
10. Lafortune EPF, Foo SC, Torrance KE, Greenberg DP (1997) Non-linear approximation of reflectance functions. In: Computer graphics (SIGGRAPH 1997), Los Angeles, pp 117–126
11. Ward GJ (1992) Measuring and modeling anisotropic reflection. In: Computer graphics (SIGGRAPH 1992). ACM, New York, pp 265–272
12. Matusik W, Pfister H, Brand M, McMillan L (2003) A data-driven reflectance model. ACM Trans Graph 22(3):759–769
13. Zickler T, Enrique S, Ramamoorthi R, Belhumeur P (2005) Reflectance sharing: image-based rendering from a sparse set of images. In: Proceedings of the eurographics symposium on rendering, eurographics association, Konstanz, pp 253–264
14. Kautz J, McCool M (1999) Interactive rendering with arbitrary BRDFs using separable approximations. In: Eurographics workshop on rendering, Granada, pp 247–260
15. Lawrence J, Rusinkiewicz S, Ramamoorthi R (2004) Efficient BRDF importance sampling using a factored representation. ACM Trans Graph 23(3):494–503
16. Rusinkiewicz S (1998) A new change of variables for efficient BRDF representation. In: Eurographics workshop on rendering, Vienna, pp 11–22
17. White DR, Saunders P, Bonsey SJ, van de Ven J, Edgar H (1998) Reflectometer for measuring the bidirectional reflectance of rough surfaces. Appl Opt 37(1):3450–3454
18. Li H, Foo SC, Torrance KE, Westin SH (2006) Automated three-axis gonioreflectometer for computer graphics applications. Opt Eng 45:043605
19. Ngan A, Durand F, Matusik W (2005) Experimental analysis of BRDF models. In: Proceedings of the eurographics symposium on rendering, eurographics association, Konstanz, pp 117–226
20. Ghosh A, Achutha S, Heidrich W, O'Toole M (2007) BRDF acquisition with basis illumination. In: IEEE 11th international conference on computer vision, 2007 (ICCV 2007), Rio de Janeiro, pp 1–8
21. Dana KJ, Wang J (2004) Device for convenient measurement of spatially varying bidirectional reflectance. J Opt Soc Am A 21(1):1–12
22. Marschner S, Westin S, Lafortune E, Torrance K, Greenberg D (1999) Image-based BRDF measurement including human skin. In: Eurographics rendering workshop 99, Granada, pp 139–152
23. Bayer BE (1976) Color imaging array. United States Patent No. 3971065
24. Lensch HPA, Kautz J, Goesele M, Heidrich W, Seidel HP (2001) Image-based reconstruction of spatially varying materials. In: Eurographics workshop on rendering, London, pp 63–70
25. Goldman DB, Curless B, Hertzmann A, Seitz SM (2005) Shape and spatially-varying BRDFs from photometric stereo. In: IEEE international conference on computer vision, Beijing
26. Woodham RJ (1980) Photometric method for determining surface orientation from multiple images. Opt Eng 19(1):139–144
27. Alldrin N, Zickler T, Kriegman D (2008) Photometric stereo with non-parametric and spatially-varying reflectance. In: IEEE conference on computer vision and pattern recognition (CVPR). IEEE, Piscataway
28. Wang J, Zhao S, Tong X, Snyder J, Guo B (2008) Modeling anisotropic surface reflectance with example-based microfacet synthesis. ACM Trans Graph 27(3):41
29. Dana K, van Ginneken B, Nayar S, Koenderink J (1999) Reflectance and texture of real-world surfaces. ACM Trans Graph 18(1):1–34
30. Han JY, Perlin K (2003) Measuring bidirectional texture reflectance with a kaleidoscope. ACM Trans Graph 22(3):741–748

Reference Plane

Zhengyou Zhang
Microsoft Research, Redmond, WA, USA

Related Concepts

►Camera Calibration; ►Perspective Camera; ►Perspective Transformation; ►Projection Transformation

Definition

A reference plane is a plane used as a reference or a constraint to solve various computer vision problems.

Background

A reference plane, because of its known geometry, can be used in many applications:
- Plane-based camera calibration [1]
- Single-view metrology [2]
- Ground plane as a reference plane for vehicle navigation
- 3D structure recovery using a reference plane

The key observation is that an image of a plane is related to the plane in space by a homography (plane projective transformation). A point not on the plane will not be mapped by the homography, resulting the so-called plane parallax.

Theory

Projective transformation is a concept used in projective geometry to describe how a set of geometric objects maps to another set of geometric objects in *projective space*. The basic intuition behind projective space is to add extra points (points at infinity) to Euclidean space, and the geometric transformation allows to move those extra points to traditional points, and vice versa.

Homogeneous coordinates are used in projective space much as Cartesian coordinates are used in Euclidean space. A point in two dimensions is described by a 3D vector. A point in three dimension is described by a 4D vector. If the homogeneous coordinates of a given point are multiplied by a nonzero scalar, the resulting homogeneous coordinates represent the same point. That is, $\lambda \boldsymbol{m}$ ($\lambda \neq 0$) and $\boldsymbol{m}$ represent the same point. Consider a point $\boldsymbol{p} = [u, v]^T$ on a plane in Euclidean space; its corresponding homogeneous coordinates are $\boldsymbol{m} = \lambda[u, v, 1]^T$. A point at infinity on the plane is represented by $[\alpha, \beta, 0]^T$, i.e., the last element is 0. A point at infinity in 2D space can be used to describe the direction of a line on the plane. Now consider a point $\boldsymbol{p} = [x, y, z]^T$ in 3D Euclidean space; its corresponding homogeneous coordinates are $\boldsymbol{m} = \lambda[x, y, z, 1]^T$. A point at infinity in 3D space is represented by $[\alpha, \beta, \gamma, 0]^T$, i.e., the last element is 0.

Projective linear transformations do not preserve sizes and angles. They do preserve incidence (e.g., points on a line remain on a line after transformation; two lines intersecting with each other will intersect after transformation) and cross ratio. A projective linear transformation are also known as a *collineation* or *projectivity*. In the case of projective plane ($\mathcal{P}^2$), it is also known as a *homography* or *plane projectivity*. In computer vision, homography plays an important role because any two images of the same planar surface are related by a homography.

A camera is modeled by the usual pinhole (see Fig. 1). A 2D point is denoted by $\boldsymbol{m} = [u, v]^T$. A 3D point is denoted by $\boldsymbol{M} = [X, Y, Z]^T$. We use $\widetilde{\boldsymbol{x}}$ to denote the augmented vector by adding 1 as the last element: $\tilde{\mathbf{m}} = [u, v, 1]^T$ and $\tilde{\mathbf{M}} = [X, Y, Z, 1]^T$. The relationship between the 3D point $\boldsymbol{M}$ and its image projection $\boldsymbol{m}$ is given by

$$s\tilde{\mathbf{m}} = \mathbf{A}\begin{bmatrix}\mathbf{R} & \boldsymbol{t}\end{bmatrix}\tilde{\mathbf{M}} \tag{1}$$

where s is a scale factor; $(\mathbf{R}, \boldsymbol{t})$, called the extrinsic parameters, is the rotation and translation which relates the world coordinate system to the camera coordinate system; and $\mathbf{A}$ is called the camera intrinsic matrix.

Without loss of generality, we assume the reference plane is on $Z = 0$ of the world coordinate system. Let us denote the i^{th} column of the rotation matrix $\mathbf{R}$ by $\boldsymbol{r}_i$. From Eq. (1), we have

$$s\begin{bmatrix}u\\v\\1\end{bmatrix} = \mathbf{A}\begin{bmatrix}\boldsymbol{r}_1 & \boldsymbol{r}_2 & \boldsymbol{r}_3 & \boldsymbol{t}\end{bmatrix}\begin{bmatrix}X\\Y\\0\\1\end{bmatrix} = \mathbf{A}\begin{bmatrix}\boldsymbol{r}_1 & \boldsymbol{r}_2 & \boldsymbol{t}\end{bmatrix}\begin{bmatrix}X\\Y\\1\end{bmatrix}.$$

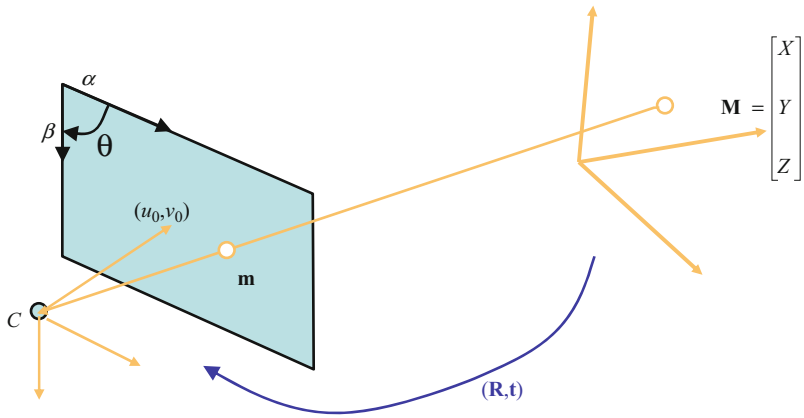
By abuse of notation, we still use $\boldsymbol{M}$ to denote a point on the model plane, but $\boldsymbol{M} = [X, Y]^T$ since Z is always equal to 0. In turn, $\tilde{\mathbf{M}} = [X, Y, 1]^T$. Therefore, a model point $\boldsymbol{M}$ and its image $\boldsymbol{m}$ is related by a homography $\mathbf{H}$:

$$s\tilde{\mathbf{m}} = \mathbf{H}\tilde{\mathbf{M}} \qquad \text{with} \quad \mathbf{H} = \mathbf{A}\begin{bmatrix}\boldsymbol{r}_1 & \boldsymbol{r}_2 & \boldsymbol{t}\end{bmatrix}. \tag{2}$$

As is clear, the 3×3 matrix $\mathbf{H}$ is defined up to a scale factor.

Furthermore, two images of the same plane are related with each other also by a homography. We use

Reference Plane, Fig. 1
Pinhole camera model



the superscript $^{(1)}$ and $^{(2)}$ to indicate the image points related to images 1 and 2, respectively. From Eq. (2), we have

$$s^{(1)}\tilde{\mathbf{m}}^{(1)} = \mathbf{H}^{(1)}\tilde{\mathbf{M}}\,, \tag{3}$$

$$s^{(2)}\tilde{\mathbf{m}}^{(2)} = \mathbf{H}^{(2)}\tilde{\mathbf{M}}\,. \tag{4}$$

It is easy to see that image point $\boldsymbol{m}^{(2)}$ is related to image point $\boldsymbol{m}^{(1)}$ by

$$s^{(21)}\tilde{\mathbf{m}}^{(2)} = \mathbf{H}^{(21)}\tilde{\mathbf{m}}^{(1)} \quad \text{with } \mathbf{H}^{(21)} = \mathbf{H}^{(2)}\mathbf{H}^{(1)-1}\,. \tag{5}$$

$\mathbf{H}^{(21)}$ is the homography from images 1 to 2.

For a point not on the reference, say $\boldsymbol{P} = [X, Y, Z, 1]^T$, its projection onto the reference plane is $\boldsymbol{M} = [X, Y]$. Mapping it to the image plane with the homography gives

$$s\tilde{\mathbf{m}} = \mathbf{H}\tilde{\mathbf{M}} = \mathbf{A}\begin{bmatrix}\boldsymbol{r}_1 & \boldsymbol{r}_2 & \boldsymbol{t}\end{bmatrix}\tilde{\mathbf{M}}\,.$$

The real image of point $\boldsymbol{P}$, however, is given by

$$t\tilde{\mathbf{p}} = \mathbf{A}\begin{bmatrix}\mathbf{R} & \boldsymbol{t}\end{bmatrix}\tilde{\mathbf{P}} = \mathbf{A}\begin{bmatrix}\boldsymbol{r}_1 & \boldsymbol{r}_2 & \boldsymbol{t}\end{bmatrix}\tilde{\mathbf{M}} + \mathbf{A}\boldsymbol{r}_3 Z\,.$$

Clearly, the plane-mapped point $\boldsymbol{m}$ is not the same as the real image point $\boldsymbol{p}$. The difference is the plane parallax. The farther away the point $\boldsymbol{P}$ is from the reference plane, the larger the plane parallax is. Plane parallax is an important quantity in 3D structure recovery based on a reference plane.

References

1. Zhang Z (2000) A flexible new technique for camera calibration. IEEE Trans Pattern Anal Mach Intell 22(11):1330–1334
2. Criminisi A (2001) Accurate visual metrology from single and multiple uncalibrated images. Springer, London/New York

Reflectance Map

Robert J. Woodham
Department of Computer Science, University of British Columbia, Vancouver, BC, Canada

Related Concepts

▸Photometric Stereo; ▸Radiance

Definition

A reflectance map is a function that gives scene radiance as a function of surface orientation.

Background

The amount of light reflected by a surface element in a given direction depends on its optical properties, on its microstructure, and on the spatial and spectral distribution of the light sources. For many materials, the fraction of the total irradiance reflected toward the viewer depends only on the surface orientation. Horn [1] introduced the reflectance map as a way to

specify scene radiance as a function of surface orientation, given the following simplifying assumptions:

1. The direction toward the viewer is the same at every visible point on the surface. This holds when image projection is orthographic.
2. The direction toward the light source is the same at every visible point on the surface. This holds for point light sources at infinity and for parallel (i.e., collimated) light sources.
3. Reflectance is isotropic about the surface normal. This holds when there is no inherent directionality in surface microstructure, making reflectance invariant to rotation of the surface about the surface normal.

A reflectance map compiles the relevant information about surface material, light source distribution, and viewing geometry into a single function. It is central to *physics-based vision* including *shape from shading* and *photometric stereo*.

Work to recover height profiles from intensity measurements originated in lunar astronomy. The height variation of points on the lunar surface is small compared to the overall distance of the moon from the earth, making image projection effectively orthographic. Direct sun illumination on the lunar surface is effectively illumination of a point light source at infinity.

As originally formulated by Horn [1], the reflectance map used the gradient to represent surface orientation. Earlier, Mackworth [2] used gradient space to express geometric constraints in the interpretation of line drawings of polyhedra. Defining the reflectance map as a function of the gradient combines radiometric and geometric constraints in a common representation, the gradient space. Subsequent variants of the reflectance map have used other representations for surface orientation including the (unit) surface normal, spherical coordinates, and stereographic coordinates.

Theory

A standard geometry is assumed. Let the visible surface be given explicitly by $z = f(x, y)$ in a left-handed Euclidean coordinate system, where the viewer is looking in the positive z direction, image projection is orthographic, illumination is parallel, and the image in the xy-axes coincides with the scene in the xy-axes.

The surface *gradient* (p, q) is defined by

$$p = \frac{\partial f(x, y)}{\partial x} \quad \text{and} \quad q = \frac{\partial f(x, y)}{\partial y} \tag{1}$$

so that a surface normal vector is $[p, q, -1]$. The gradient (p, q) is one way to represent the two degrees of freedom of surface orientation.

The reflectance map, $R(p, q)$, determines scene radiance as a function of the gradient for a specific surface material, scene illumination, and viewing geometry. Further, if an ideal (calibrated) camera produces image intensity proportional to scene radiance, then the image irradiance equation is

$$E(x, y) = R(p, q) \tag{2}$$

where $E(x, y)$ is image irradiance. Equation (2) is a nonlinear, first-order partial differential equation. *Shape from shading* methods determine a surface, $z = f(x, y)$, that satisfies the image irradiance equation over some domain, Ω, including any initial conditions specified on the boundary, $\partial\Omega$, or elsewhere. Sometimes, a priori constraints on the reflectance map simplify shape analysis. Three such cases are:

1. If $R(p, q)$ is symmetric about the origin in gradient space, then it is a function of $\sqrt{p^2 + q^2}$ alone and Eq. (2) is eikonal. An eikonal image irradiance equation can often be achieved by aligning a single light source direction with the viewing direction.
2. For the special case of material in the maria of the moon and other materials for which reflectance depends only on the ratio of the incident and emergent angles, $R(p, q)$ is linear in p and q.
3. For an ideal diffuse (Lambertian) reflector, Eq. (2) becomes linear if the (unit) surface normal, rather than the gradient, is used to represent surface orientation.

Application

Generic robot vision tasks such as object recognition, pose determination, and inspection typically assume that measured brightness depends upon surface shape. When the illumination and surface material are fixed, it becomes possible to relate measured brightness directly to shape, as the reflectance map demonstrates.

Remote sensing, on the other hand, typically assumes that (multispectral) measurements define a "spectral signature" that depends upon surface material (i.e., ground cover). Not surprisingly, difficulties arise when rugged terrain and illumination change confound the measurements. Reflectance maps have been used to decouple geometric effects, associated with elevation, slope, and aspect, and from spectral effects, associated with surface material.

Determining the Reflectance Map

Reflectance maps are used in methods to determine shape and surface material from measured brightness and color. But, how are the appropriate reflectance maps determined? There are three approaches:

1. Reflectance can be modeled phenomenologically. That is, one imagines how an idealized material might reflect light and derives the expression for scene radiance accordingly. Ideal diffuse (Lambertian) reflection is one example of a phenomenological model. Phong shading, popular in computer graphics, is another. A reflectance map is obtained when the gradient (p, q) is used to represent surface orientation.
2. There is a standard nomenclature for reflectance [3]. The intrinsic reflectance properties of a surface material are specified by its *bidirectional reflectance distribution function* (BRDF). When the BRDF and light source distribution are known, the reflectance map can be derived analytically. Specific examples are given in [4].

   Analytic modeling extends to taking simple phenomenological models, like ideal diffuse or ideal specular reflection, and applying them to a surface microstructure of known (or assumed) particle shape and distribution. For example, in reflectance spectroscopy, a technique in analytic chemistry, materials are grounded into fine powders of known particle size and shape. Analytic models are developed to relate measured reflectance of the powders to the optical properties of the material of which they are composed.
3. Finally, a reflectance map can be measured directly using a goniometer-mounted sample or indirectly from images of a calibration object of known shape, such as a sphere. Empirical measurement has the benefit of automatically compensating for the transfer characteristics of the camera (or other sensor). Calibration results are directly applicable to the analysis of other objects of different shape but made of the same material as the calibration object and illuminated and viewed under the same conditions. In this way, a material with any reflectance characteristic can be handled, provided that the necessary calibration can be done. In some applications, it has been useful to use paint (or other coating) to match reflectance properties between a calibration object and other objects to be analyzed.

   Empirical measurement over a wide range of viewpoints and illumination conditions is both data and time intensive. This has led to approaches that are best termed semiempirical. Reflectance is assumed to take on a particular functional form, typically a linear combination of certain basis reflectance functions, and empirical measurement is used to estimate the parameters associated with the assumed functional form.

Open Problems

Image irradiance equations can be generalized to accommodate perspective projection, nearby light sources, reflectance that is not isotropic about the surface normal, and optical properties of the medium through which the radiant energy is transmitted. Increasingly, the spectral dependence of reflectance also is made explicit. This adds complexity to the associated optical equations, in the case of analytic models, and to the associated calibration/storage requirements, in the case of empirical models.

Unfortunately, with empirical models, it is difficult to use measurements acquired under one condition of illumination and viewing to predict the reflectance map for other conditions of illumination and viewing. Specifically, there is no fundamental, scale-independent distinction to be made among intrinsic optical properties (i.e., the BRDF), surface microstructure, and gross surface shape.

Any reflectance map is subject to error in the presence of cast shadows and interreflection. No purely local definition can succeed since these phenomena are inherently nonlocal. Interreflection, for example, causes changes in the local illumination owing to the secondary reflections from adjacent object points.

References

1. Horn BKP (1977) Understanding image intensities. Artif Intell 8:201–231
2. Mackworth AK (1973) Interpreting pictures of polyhedral scenes. Artif Intell 4(2):121–137
3. Nicodemus FE, Richmond JC, Hsia JJ, Ginsberg IW, Limperis T (1977) Geometrical considerations and nomenclature for reflectance. NBS monograph, vol 160. U.S. Department of Commerce, Washington DC
4. Horn BKP, Sjoberg RW (1979) Calculating the reflectance map. Appl Opt 18:1770–1779

Reflectance Models

Szymon Rusinkiewicz
Department of Computer Science, Princeton University, Princeton, NJ, USA

Synonyms

Analytic reflectance functions; BRDF models

Definition

Reflectance models are analytic functions specifying the ratio of reflected radiance to irradiance at a point.

Background

Whether for recognition or image synthesis, it is frequently necessary to have a model for how much light a surface will reflect. Though a surface may reflect a different amount of light at each position, the bidirectional reflectance distribution function (BRDF) embodies the reflectance at a single point, for every possible angle of incident and exitant light. It is defined [1] as the ratio of reflected radiance in a direction ω_o to irradiance from direction ω_i:

$$f_r(\omega_i \to \omega_o) = f_r(\theta_i, \varphi_i, \theta_o, \varphi_o) = \frac{dL_o(\omega_o)}{dE_i(\omega_i)}. \quad [\mathrm{sr}^{-1}] \tag{1}$$

Note that this equation is defined in terms of *irradiance*, which implicitly includes the famous "cosine law" of incident light. Thus, the BRDF must be multiplied by $\cos\theta_i$ in order to obtain the full variation of reflected light as a function of the angle of incidence (for a distant light source).

Though research into photorealistic image synthesis increasingly uses dense BRDF measurements (including thousands of samples across different angles of incidence and exitance) [2], the reflectance of many surfaces is adequately predicted by simple analytic functions with few free parameters. Assuming such an analytic model makes it practical to infer the complete bidirectional reflectance from a small number of measurements, fitting the function and obtaining its parameters based on the available data [3]. This function, in turn, may be used to interpret further images.

Theory

Three general classes of reflectance models have been developed in the literature. *Physically based* models attempt to model reflectance from first principles, beginning with the solution to Maxwell's equations on surfaces of known geometry. *Microfacet* models assume a rough surface geometry, which is not known exactly but may be characterized statistically. Finally, ad hoc, phenomenological, or *empirical* models capture some qualitative features of reflectance without necessarily striving for, or achieving, certain aspects of *physically plausible* reflectance.

Physical plausibility in reflectance typically refers to two specific properties satisfied by all actual surfaces. The first is *energy conservation*: because all incident light must be either reflected or absorbed and no light may be created during reflection, it is impossible for a surface to reflect more light than was incident on it. Mathematically, the integral of the BRDF over all outgoing directions, scaled by a cosine term to account for foreshortening, must be less than one:

$$\forall \omega_i : \int_\Omega f_r(\omega_i, \omega_o) \cos\theta_o \, d\omega_o \le 1. \tag{2}$$

A second, more subtle, property of BRDFs is that they must be unchanged when the angles of incidence and exitance are swapped:

$$f_r(\omega_i \to \omega_o) = f_r(\omega_o \to \omega_i). \tag{3}$$

This is a condition known as *Helmholtz reciprocity* and is due to the symmetry of light transport [4]. Some systems, such as the work on Helmholtz stereopsis [5],

have relied on this property, which often expressed as camera/projector duality: in many imaging systems, it is possible to interchange the roles of camera and projector, provided that cosine terms are properly accounted for.

Lambertian Reflectance

We now turn to specific reflectance models. The simplest is just a constant:

$$f_\mathrm{r} = const. = \rho/\pi. \tag{4}$$

(It is important to keep in mind that the BRDF is defined in terms of irradiance, which has the "incident cosine law" implicitly included.) This results in a matte or diffuse appearance and is known as ideal Lambertian reflectance. This BRDF is frequently written as a constant ρ divided by π. In this case, ρ is interpreted as the diffuse albedo: it is the fraction of light that is reflected (vs. absorbed) by the surface, and a surface with this reflectance conserves energy precisely when the albedo is less than or equal to one. Because this model is independent of the directions of incidence and exitance, it also satisfies reciprocity.

Phong and Blinn-Phong BRDFs

Another simple analytic BRDF, designed to empirically mimic the appearance of glossy (also called shiny or specular) materials, is the Phong model [6]:

$$f_\mathrm{r} = k_\mathrm{s} (\boldsymbol{r} \cdot \boldsymbol{v})^n, \tag{5}$$

where $\boldsymbol{v}$ is the direction toward the viewer and $\boldsymbol{r}$ is the mirror reflection of the light direction from the tangent plane. Note that a frequently used version of the Phong "BRDF" includes an additional $1/\cos\theta_\mathrm{i}$ factor, which is canceled by the irradiance cosine law. The latter is therefore not a physically plausible BRDF: it does not exhibit reciprocity and does not conserve energy.

A common variant of this model is sometimes known as the Blinn-Phong model [7]:

$$f_\mathrm{r} = k_\mathrm{s} (\boldsymbol{n} \cdot \boldsymbol{h})^n, \tag{6}$$

though again it is often stated as a physically implausible shading model rather than a BRDF. In this equation, $\boldsymbol{h}$ is the "halfway" vector, which is midway between the light direction $\boldsymbol{l}$ and the viewer direction $\boldsymbol{v}$:

$$\boldsymbol{h} = \frac{\boldsymbol{l} + \boldsymbol{v}}{\|\boldsymbol{l} + \boldsymbol{v}\|}. \tag{7}$$

In contrast to the Lambertian BRDF, the distribution of reflected light in these models is not constant. In fact, there is a lobe centered around the direction of ideal mirror reflection for each incident angle, containing significantly more energy than the rest of the domain. This is known as the specular lobe, and its size and width (fall off) are controlled by the parameters k_s and n, respectively.

Lafortune BRDF

A popular model used for fitting analytic functions to measured BRDF data is the Lafortune model [8]:

$$f_\mathrm{r} = \left(C_\mathrm{x} l_\mathrm{x} v_\mathrm{x} + C_\mathrm{y} l_\mathrm{y} v_\mathrm{y} + C_\mathrm{z} l_\mathrm{z} v_\mathrm{z}\right)^n, \tag{8}$$

in which l_x, v_x, etc. represent the components of the light vector $\boldsymbol{l}$ and view vector $\boldsymbol{v}$, in a coordinate system in which the surface normal is oriented along the z-axis. This model reduces to Phong by choosing $-C_x = -C_y = C_z = \sqrt[n]{k_s}$ but through suitable choice of parameters can also represent non-Lambertian diffuse reflection, off-specular reflection, anisotropy, and retro reflection. It is also common to fit a sum of multiple lobes of (Eq. 8) to measured datasets.

Ward BRDF

Another popular BRDF used in fits to measurements is the Ward model [9]:

$$f_r = k_s \frac{\mathrm{e}^{-\tan^2\theta_\mathrm{h}\left((\cos^2\phi_\mathrm{h})/\alpha_\mathrm{x}^2 + (\sin^2\phi_\mathrm{h})/\alpha_\mathrm{y}^2\right)}}{4\pi\,\alpha_\mathrm{x}\alpha_\mathrm{y}\sqrt{\cos\theta_\mathrm{i}\cos\theta_\mathrm{o}}}. \tag{9}$$

Compared to the Blinn-Phong BRDF, the Ward model includes a specular peak shaped by a Gaussian function (as opposed to a power-of-cosine model) but also can model anisotropic reflection by using separate Gaussian widths α_x and α_y in two perpendicular directions.

Microfacet BRDFs

Numerous BRDFs have been derived from first principles that predict the aggregate reflectance for surfaces that at a small scale consist of tiny, mirror-reflective "microfacets" oriented in random directions. An early microfacet BRDF was originally developed in the physics community by Torrance and Sparrow [10],

introduced in graphics by Blinn [7], and later refined by Cook and Torrance [11]:

$$f_r = \frac{D\ G\ F}{\pi\ \cos\theta_i\ \cos\theta_o}. \tag{10}$$

There are three major terms in the model that describe the angular distribution of microfacets, how many are visible from each angle, and how light reflects from each facet.

The first term D in the Torrance-Sparrow model describes the density of facets facing in any possible direction:

$$D = \frac{e^{-(\tan^2\theta_h)/m^2}}{4m^2\ \cos^4\theta_h}, \tag{11}$$

where θ_h is the angle between the halfway vector $\boldsymbol{h}$ and the surface normal $\boldsymbol{n}$. Notice that part of this term resembles a Gaussian, and this is not a coincidence: the Torrance-Sparrow model makes the assumption that the microfacet normals have a Gaussian distribution controlled by a "roughness" parameter m. The $\cos^4\theta_h$ term occurring here is a change-of-basis term: it is included to properly normalize a probability distribution expressed in terms of the halfway vector.

The next term G in the Torrance-Sparrow model accounts for the fact that not all facets are visible from all directions, because they are hidden by the facets in front of them. It includes both "shadowing" and "masking" effects, representing occlusion from the point of view of the light and viewer, respectively:

$$G = \min\left\{1,\ \frac{2\cos\theta_h\cos\theta_i}{\cos\theta_d},\ \frac{2\cos\theta_h\cos\theta_o}{\cos\theta_d}\right\}. \tag{12}$$

This formula is derived by considering a particular microgeometry: the microfacets are assumed to form V-shaped grooves in the surface, which are symmetric about the (macroscopic) surface normal.

Finally, the reflection from each facet is described by the Fresnel term F, which predicts that reflection increases toward grazing angles. This term arises from a solution to Maxwell's equations on a surface:

$$F = \frac{1}{2}\left(F_\perp + F_\parallel\right) = \frac{1}{2}\left[\left(\frac{\sin(\theta_t - \theta_d)}{\sin(\theta_t + \theta_d)}\right)^2 + \left(\frac{\tan(\theta_d - \theta_t)}{\tan(\theta_d + \theta_t)}\right)^2\right], \tag{13}$$

where θ_d is half the angle between the incident and exitant directions, $\theta_t = \sin^{-1}\left((\sin\theta_d)/\eta\right)$, η is the index of refraction of the surface, and the two terms represent the portion of reflected light polarized perpendicular and parallel to the plane of incidence. Note that the "difference angle" θ_d is the angle of incidence (and exitance) on a microfacet oriented to produce mirror reflection between the desired angles of incidence and reflection.

More recently, Ashikhmin et al. [12] generalized these types of microfacet BRDFs to allow expressing arbitrary half angle distributions. They demonstrate how to modify these BRDFs to replace the analytic distribution in Eq. 11 with alternative analytic forms or tabulated (sampled) functions that can express arbitrary patterns.

More Complex Analytic BRDFs

In addition to models for primarily specular surfaces, physically based BRDFs have been derived for rough diffuse surfaces (the Oren-Nayar model [13]) and for dusty surfaces (the Hapke/Lommel-Seeliger model, developed to model lunar reflectance [14]). They range in complexity from simple formulas that ignore many real-world effects to complex models that attempt to account for most actually observed surface phenomena (e.g., the He-Torrance-Sillion-Greenberg model [15]). While a detailed description of these models is beyond the scope of this entry, they are sometimes used in rendering or vision systems. The Oren-Nayar model, in particular, is often combined with the Torrance-Sparrow model (with the abbreviation TSON) to model surfaces with both a specular and non-Lambertian diffuse component. One drawback of these models, however, is that their additional complexity and many parameters can make it difficult or unstable to fit them to measured data.

References

1. Nicodemus FE, Richmond JC, Hsia JJ, Ginsberg IW, Limperis T (1977) Geometric considerations and nomenclature for reflectance. Monograph 160. National Bureau of Standards (US)
2. Matusik W, Pfister H, Brand M, McMillan L (2003) A data-driven reflectance model. ACM Trans Graph (SIGGRAPH 2003) 22(3):759–770

3. Ngan A, Durand F, Matusik W (2005) Experimental analysis of brdf models. In: Proceedings of the 15th eurographics symposium on rendering, eurographics association, pp 117–226
4. von Helmholtz H (1867) Handbuch der Physiologischen Optik. Leopold Voss, Leipzig
5. Zickler T, Belhumeur PN, Kriegman DJ (2002) Helmholtz stereopsis: exploiting reciprocity for surface reconstruction. In: ECCV '02: proceedings of the 7th european conference on computer vision-part III (ECCV). Springer, London, pp 869–884
6. Phong BT (1975) Illumination for computer generated pictures. Commun ACM 18(6):311–317
7. Blinn JF (1977) Models of light reflection for computer synthesized pictures. In: SIGGRAPH '77: proceedings of the 4th annual conference on computer graphics and interactive techniques. ACM, New York, pp 192–198
8. Lafortune EPF, Foo SC, Torrance KE, Greenberg DP (1997) Non-linear approximation of reflectance functions. In: SIGGRAPH '97: proceedings of the 24th annual conference on computer graphics and interactive techniques. ACM/Addison-Wesley, New York, pp 117–126
9. Ward G (1992) Measuring and modeling anisotropic reflection. Comput Graph 26(Annual Conference Series): 265–273
10. Torrance KE, Sparrow EM (1967) Theory for off-specular reflection from roughened surfaces. JOSA 57(9): 1105–1114
11. Cook RL, Torrance KE (1982) A reflection model for computer graphics. ACM Trans Graph 1(1):7–24
12. Ashikhmin M, Premoze S, Shirley P (2001) A microfacet-based BRDF generator. In: Computer graphics (Proceedings of ACM SIGGRAPH 2001), pp 65–74
13. Oren M, Nayar S (1994) Generalization of lambert's reflectance model. In: Proceedings of SIGGRAPH
14. Hapke B (1963) A theoretical photometric function for the lunar surface. J Geophys Res 68(15)
15. He XD, Torrance KE, Sillion FX, Greenberg DP (1991) A comprehensive physical model for light reflection. Comput Graph 25(Annual Conference Series):175–186

Reflection Mapping

▸Image-Based Lighting

Reflectometry

▸Recovery of Reflectance Properties

Relation Between Objects and Their Digital Images

▸Digitization

Relief Texture

▸Bidirectional Texture Function and 3D Texture

Retina

▸Image Plane

Retinex Algorithm

▸Retinex Theory

Retinex Theory

Micah K. Johnson
Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA

Synonyms

Retinex algorithm

Related Concepts

▸Color Constancy; ▸White Balance

Definition

Retinex theory is a computational model for human color constancy. It defines a mechanism for computing lightness values from an image. Retinex theory proposes that the lightness values for each class of photoreceptors are derived independently and that this triplet of values correlates with perceived reflectance.

Background

The human visual system is remarkable in its ability to deal with varying illumination. Throughout the day, the visual system encounters both artificial and natural

Retinex Theory, Fig. 1 An arrangement of colored rectangles similar to "Mondrian-like" stimuli used by Land and McCann [1]

illumination, yet the appearance of the world seems stable. For example, a piece of white paper on a desk does not appear to change color when taken outside.

But anyone who has used a camera knows that the world is not as simple as it appears. Sometimes photographs captured outdoors have a blue color cast, while photographs captured indoors appear yellow. These color shifts are due to the different sources of illumination, e.g., sunlight versus an incandescent bulb. For a camera to capture the scene as it appears to the human visual system, it needs to determine, or be configured with, the color of the illumination. This setting in the camera is known as white balance.

While a camera needs to be configured for the type of illumination, the human visual system accounts for illumination changes automatically. This ability is known as color constancy. The process is fast and effortless – many people do not realize how much the measured color of an object changes under different light sources because its appearance seems stable. In a sense, the human visual system has an automatic white balance mechanism that operates without the need for a reference or knowledge of the illumination.

Land and McCann investigated color constancy under varying illumination in a series of experiments known as the "Mondrian" experiments [1]. The stimuli for the experiments consisted of a large array of rectangular colored papers that were arranged to resemble a painting by the artist Mondrian; see Fig. 1. The papers were illuminated by three projectors with filters designed to pass long waves, middle-length waves, or short waves. Focusing on one of the papers (e.g., a green paper), they adjusted the flux from each of the projectors to achieve a predetermined set of luminance values at the eye, as measured by a photometer. The subject in the study noted the color name of the paper in question. The process was repeated for other pieces of paper – first adjusting the lights to achieve the same set of luminance values, then asking the subject to note the color name. The subjects consistently reported the correct color names for the papers, despite receiving the same luminance triple at the eye from all papers. Thus, in a controlled environment, Land and McCann verified the ability of the visual system to identify correct color names under varying illumination, even when the luminance values at the eye in all cases were the same.

Theory

Retinex theory grew out of the Mondrian experiments and it defined a computational model for the color constancy of the Mondrian stimuli. Land and McCann coined the name "Retinex" to signify that, in humans, this process could involve both the retina and the cortex [1].

The main principle of Retinex theory is that images from the three classes of photoreceptors (i.e., color channels) are processed separately into three *lightness* images. The lightness values at different regions in the image correlate with the true reflectance of the region and are as independent of illumination differences as possible. Therefore, the color constancy problem is cast into a lower-dimensional problem of understanding the lightness of a single-channel image.

Retinex theory describes an algorithm for computing lightness images from intensity images. There were five main components in the original algorithm: ratio, sequential product, reset, threshold, and average [1].

Land and McCann realized the perceptual importance of edges and made the ratio of luminances across edges the fundamental unit of their algorithm. To compare areas that were not directly next to each other, they introduced the notion of a sequential product.

The sequential product accumulates edge ratios along a one-dimensional path. The initial value for the sequential product is set to one, and the sequential product can become greater than one if the path crosses a region with greater reflectance than the initial region. In this case, the sequential product is reset to one to account for the highest reflectance found along the path.

Across large areas of uniform reflectance, adjacent pixels may have slightly different luminance values, leading to ratios near, but not equal to, one. The threshold component discards these small changes to make the algorithm more robust to gradual changes of illumination across the scene.

The sequential product reveals the relative reflectance of a region with respect to the highest reflectance seen along a one-dimensional path. The Retinex algorithm considers several one-dimensional paths through the same region and averages the relative reflectances to obtain the lightness value for the region.

Through the five operations, the Retinex algorithm converts a grayscale intensity image into a lightness image. The triple of lightness values from three color channels were shown to correlate with color judgments, regardless of illumination conditions [2].

Since the original Retinex paper, there have been further studies on the roles of the different operators and modifications to these operators [3, 4]. For example, the length of the path used in the sequential product affects the ability of the algorithm to model color constancy. Short paths include little of the surrounding context, while in the limit, infinite paths result in a scaling of the image and are a poor model for human perception [5]. Later variants of the Retinex algorithm also discarded the threshold step [6], though other authors have found the threshold step to be important for separating shading and reflectance effects in images [7–9].

References

1. Land EH, McCann JJ (1978) Lightness and retinex theory. J Opt Soc Am 61(1):1–11
2. Land EH (1977) The retinex theory of color vision. Sci Am 237(6):108–128
3. Land EH (1983) Recent advances in retinex theory and some implications for cortical computations: color vision and the natural image. Proc Natl Acad Sci USA 80:5163–5169
4. Land EH (1986) An alternative technique for the computation of the designator in the retinex theory of color vision. Proc Natl Acad Sci USA 83:3078–3080
5. Brainard DH, Wandell BA (1986) Analysis of the retinex theory of color vision. J Opt Soc Am A 3(10):1651–1661
6. Funt B, Ciurea F, McCann J (2004) Retinex in MATLAB. J Electron Imaging 13(1):48–57
7. Horn BKP (1974) Determining lightness from an image. Comput Graph Image Process 3:277–299
8. Blake A (1985) Boundary conditions for lightness computation in Mondrian world. Comput Vis Graph Image Process 32:314–327
9. Tappen MF, Freeman WT, Adelson EH (2005) Recovering intrinsic images from a single image. IEEE Trans Pattern Anal Mach Intell 27(9):1459–1472

Retroreflection

S. C. Pont
Industrial Design Engineering, Delft University of Technology, Delft, The Netherlands

Synonyms

Backscattering

Related Concepts

▸Asperity Scattering; ▸Lambertian Reflectance; ▸Surface Roughness

Definition

Retroreflection is a type of reflectance which is characterized by a peak in the backward or illumination direction. Thus, materials that backscatter scatter light primarily in the direction from which it is illuminated.

Background

The reflectance of natural, opaque, rough surfaces can be described by the Bidirectional Reflectance Distribution Function (BRDF) [7]. BRDFs that are common and well known are those of Lambertian, perfectly diffusely scattering, and of specular surfaces. Such surfaces scatter light in all directions (diffuse scattering) or primarily in the mirror direction (specular reflection). However, natural surfaces can scatter light in many other ways. Retroreflection is just one such manner. Retroreflection can be caused by, for instance, hemispherical concavities in the surface [5, 8]. Even

if the surface is locally Lambertian, such a "thoroughly pitted surface" shows retroreflection. Another example is the triple mirror retroreflector, which is, for instance, used in reflectors for cars and bicycles (in small-scale versions) and in distance measurements (relatively large-scale versions). Many reflectors for outside applications (e.g., clothes of roadworkers, number plates) consist of a thin layer of small spherical particles with a refraction index of 2. Natural phenomena that are caused by retroreflection are "Heiligenschein" (the halo you can see around your head if you look at wet or dewy grass with the sun at a low angle behind you; for examples see [3]) and the Seeliger effect (reflective objects that are in opposition to the sun are brighter than in other positions).

Theory

The geometrical optical models with retroreflection lobes which were mentioned above result in quite complicated BRDFs. It is possible to fit retroreflection characteristics in a convenient, simplified formula (note that basic physical constraints should hold, e.g., nonnegativity, energy conservation, and Helmholtz reciprocity) [4]. For instance, the following BRDF model,

$$B(\mathbf{i},\mathbf{j},\mathbf{n}) = \frac{1}{\pi 2^k}\frac{(1+\mathbf{i}\cdot\mathbf{j})^k}{(\mathbf{i}+\mathbf{j})\cdot\mathbf{n}}, \qquad (1)$$

is a perfect backscatterer in the sense that the BRDF equals $(2\pi(\mathbf{i}\cdot\mathbf{n}))^{-1}$ for coincident directions of incidence and viewing ($\mathbf{i}=\mathbf{j}$): When the normal irradiance is H_N, the irradiance becomes $H_N(\mathbf{i}\cdot\mathbf{n})$ and the backscattered radiance $H_N/2\pi$, i.e., independent of the slant of the surface. Thus, the full moon would appear as a featureless disk. It may well be the simplest analytical expression that leads to a pronounced backscatter lobe. The BRDF peaks at $\mathbf{i}=\mathbf{j}$; the parameter k determines the width of the peak. The albedo is a complicated function, for instance, for $k=1$ the albedo is

$$A_B(\vartheta_i) = \frac{1}{2}(1+\cos\vartheta_i-\cos 2\vartheta_i) + \frac{1}{4}(\cos\vartheta_i-\cos 3\vartheta_i)\log\frac{\cos\vartheta_i}{1+\cos\vartheta_i}, \qquad (2)$$

where $\cos\vartheta_i = \mathbf{i}\cdot\mathbf{n}$. This albedo is $\frac{1}{2}$ for normal incidence and rises to 1 for grazing incidence. For $k>1$ the expressions for the albedo become unwieldy, but numerical integration reveals that the albedo remains lower than one and approximately constant for a large range, then rises to one at grazing incidence. It is not obvious how to write down a backscatter BRDF with unit albedo throughout. Other possibilities for simplified formulations may be found in graphics as so-called backscatter shaders. However, care should be taken that many of these rendering applications do not fulfill the above-mentioned basic physical constraints.

Open Problems

BRDFs of natural surfaces can probably be categorized into about a dozen different modes. Currently only the forward, backward, diffuse, and surface scattering modes have been described by formal optical models.

Reflectance estimation from images suffers from image ambiguities. Prior knowledge on the reflectance statistics of natural materials plus formal descriptive models for the common modes of natural BRDFs can constrain this problem.

References

1. CUReT, Columbia–Utrecht reflectance and texture database. http://www.cs.columbia.edu/CAVE/curet
2. Dana KJ, van Ginneken B (1977) Reflectance and texture of real-world surfaces. In: Proceedings of the IEEE computer science conference on computer vision and pattern recognition (CVPR)
3. Heiligenschein photographs, http://www.weatherscapes.com/album.php?cat=optics&subcat=heiligenschein
4. Koenderink JJ, Pont SC (2008) Material properties for surface rendering. Int J Comput Vis Biomech 1(1):43–53
5. Koenderink JJ, van Doorn AJ, Dana KJ, Nayar S (1999) Bidirectional reflection distribution function of thoroughly pitted surfaces. Int J Comput Vis 31(2/3):129–144
6. Lambert JH (1760) Photometria Sive de Mensure de Gradibus Luminis, Colorum et Umbræ. Eberhard Klett, Augsburg
7. Nicodemus FE, Richmond JC, Hsia JJ (1977) Geometrical considerations and nomenclature for reflectance. National Bureau of Standards, U.S. Monograph, vol 160. U.S. Department of Commerce, Washington
8. Pont SC, Koenderink JJ (2002) Bidirectional reflectance distribution function of specular surfaces with hemispherical pits. J Opt Soc Am A 19(12):2456–2466

Rigid Alignment

▸Rigid Registration

Rigid Matching

▸Rigid Registration

Rigid Positioning

▸Rigid Registration

Rigid Registration

Yonghuai Liu[1], Ralph R. Martin[2], Liang Chen[3], Xuejun Ren[4] and Longzhuang Li[5]
[1]Department of Computer Science, Aberystwyth University, Ceredigion, Wales, UK
[2]School of Computer Science and Informatics Cardiff University, Cardiff, UK
[3]Department of Computer Science, University of Northern British Columbia, Prince George, Canada
[4]School of Engineering Liverpool John Moores University, Liverpool, UK
[5]Department of Computing Science, Texas A and M University at Corpus Christi, Corpus Christi, TX, USA

Synonyms

Rigid alignment; Rigid matching; Rigid positioning; Rigid transformation estimation

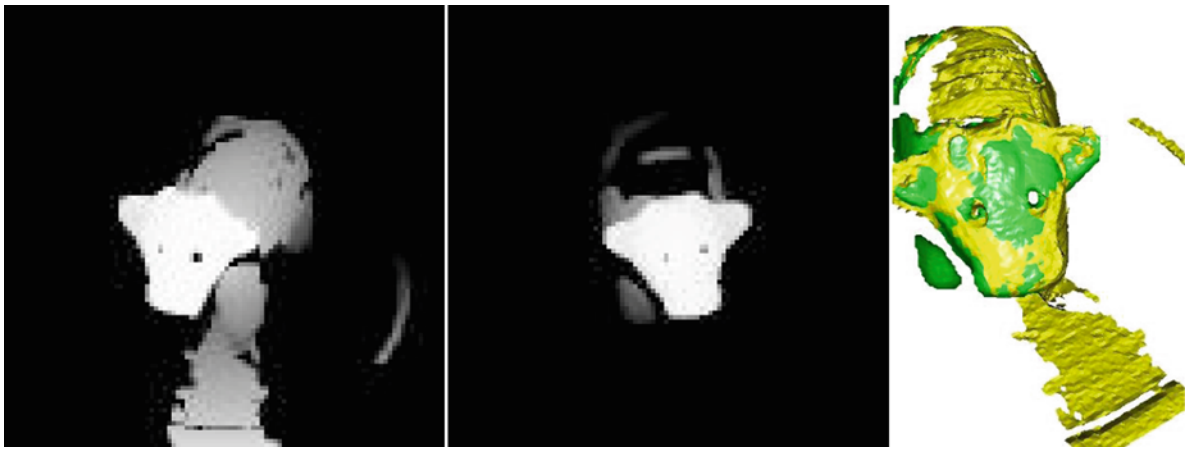
Definition

Given two copies of an object surface at different locations and orientations in space, or two parts of the surface of a single object with at least some shared overlapping area, find a translation and rigid rotation which places the objects, or corresponding parts of the object, at the same location and orientation. This process is called rigid registration. In practice many approaches to rigid registration work by finding point-to-point correspondences between parts of the object surface in each dataset and use these to estimate the geometric transformation in either least-squares or weighted least-squares sense with closed-form solution. Often, registration algorithms also output the point-to-point correspondences which can be just as useful to many applications as the transformation itself. A correspondence is such a pair of points that while they are described in two different coordinate systems, they represent the same point on the object surface in 3D space.

Sometimes, multiple captured datasets need to be registered either simultaneously or sequentially in a pairwise manner. Sometimes, captured data need to be registered against, say, a stored CAD model of an ideal version of the object, etc. Thus, in some cases, the "reference" shape comes from the data itself, while in other cases, there is an external reference shape.

Background

Current laser scanning technologies enable the acquisition of both depth and intensity information from an object of interest in the form of range (depth) and intensity images as illustrated in Fig. 1. In this case, the object of interest is a (toy) cow and is represented as structured points. In Fig. 1, holes can be seen on the head and back of the cow due to inability of the scanner to capture data: its hind legs cannot be seen, while a circle and patch underneath the cow belong to the background. The occurrences and locations of such holes, occlusions, and clutter vary from one dataset to another and are unpredictable. While the range images describe the geometry of the object of interest, the intensity images describe its appearance. Since laser scanning systems (range cameras) have a limited field of view, and one part of the object may occlude another, a number of images have to be captured from different viewpoints (Fig. 1) to obtain reasonably complete coverage of the object surface; even then small gaps may remain. Each image is recorded in a local laser scanning system-centred coordinate system. To combine these images to give an overall model of the object by fusing the geometrical and optical information in these images, all images have to be placed and aligned in a single global coordinate system. This process is

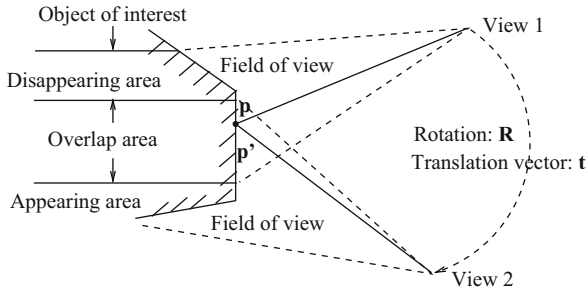


Rigid Registration, Fig. 1 Range images of a cow captured from two different viewpoints [1] are superimposed after registration. *Left*: cow45; *middle*: cow48; *right*: in the result, differently colored points belong to the two different original point sets

called registration. Shape registration has two goals: one is to establish correspondences between two overlapping shapes; the other is to estimate the underlying transformation parameters that bring one shape into the best possible alignment with the other. Fixing either of these two goals renders the other easier. However, they are in practice interwoven, complicating the shape registration process.

Suppose that the object of interest is rigid and represented using a set of points. Then interpoint Euclidean distances on the object surface are invariant with respect to changes in viewpoint from which the object is imaged. Let a general point on the object surface belong to the overlapping area between two views. Suppose it is seen in these two views as points with coordinates $(\mathbf{p}, \mathbf{p}')$. This pair is called a correspondence between the views. These coordinates are related by a rigid transformation (Fig. 2), which may be expressed using a rigid rotation matrix $\mathbf{R}$ and a translation vector $\mathbf{t}$: $\mathbf{p}' = \mathbf{R}\mathbf{p} + \mathbf{t}$. As a rigid rotation matrix, $\mathbf{R}$ has to satisfy two conditions: (1) unit determinant: $\det(\mathbf{R}) = 1$ and (2) orthonormality: $\mathbf{R}^T\mathbf{R} = \mathbf{I}$. $\mathbf{R}$ and $\mathbf{t}$ describe the relative orientation and position, respectively, of points seen in the second scene in the coordinate system of the first scene.

Automatic shape registration is a challenging problem for various reasons: (1) imaging noise: the captured data is corrupted by noise due to spatial sampling, variations in reflectance properties, depth discontinuities, mechanical control errors, signal



Rigid Registration, Fig. 2 The relationship of a correspondence $(\mathbf{p}, \mathbf{p}')$ between two views $\mathbf{V}_1$ and $\mathbf{V}_2$ can be described using a rigid rotation matrix $\mathbf{R}$ and a translation vector $\mathbf{t}$: $\mathbf{p}' = \mathbf{R}\mathbf{p} + \mathbf{t}$

quantization, signal detection processes, etc. (2) occlusion: points may be absent in certain views, as one part of the object hides another; (3) limited field of view: this characteristic of scanning devices means that there is often relatively little overlap between adjacent views; (4) ambiguity between foreground and background: to facilitate data capture, a background is usually set up with different reflectance properties from the foreground to help separate background data from the foreground data. However, this is not always fully successful, resulting in foreground data being contaminated by background data; and (5) computational complexity: the amount of captured data is large, leading to ever more time-consuming computations as scanner resolution and field of view increase.

Theory

There are many approaches to registration such as feature extraction and matching [2], an optimal match of points in different shapes [3, 4], and randomized transformation search [5]. The iterative closest point (ICP) algorithm [4] was the first popular approach, and there are many variants [6–9] of it. Here a state-of-the-art approach is proposed. Assume that the two shapes to be registered are represented as two sets of unorganized points $\mathbf{P} = \{\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_{n_1}\}$ and $\mathbf{P}' = \{\mathbf{p}'_1, \mathbf{p}'_2, \cdots, \mathbf{p}'_{n_2}\}$, representing two different *parts* of the same free-form shape from two different nearby viewpoints with a certain amount of overlap in 3D space (Fig. 2). Suppose that the underlying transformation parameters rotation matrix $\mathbf{R}$ and translation vector $\mathbf{t}$ have somehow been initialized or estimated. Then a weight w_{ij} is defined for the representation of the extent to which the point pair $(\mathbf{p}_i, \mathbf{p}'_j)$ represents a correct tentative correspondence between $\mathbf{P}$ and $\mathbf{P}'$. The definition for w_{ij} must take into account imaging noise, appearance and disappearance of points, and clutter in both $\mathbf{P}$ and $\mathbf{P}'$. To estimate w_{ij}, the following objective function is constructed:

$$J(\mathbf{W}) = \text{argmin}_{\mathbf{W}} \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} w_{ij} ||\mathbf{p}'_j - \mathbf{R}\mathbf{p}_i - \mathbf{t}||^2 - \left(-\frac{1}{\beta} \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} w_{ij} (\ln w_{ij} - 1) \right) \quad (1)$$

where $\mathbf{W} = \{w_{ij}\}$. β is a positive parameter used to describe the extent to which it is believed that the transformation parameters $\mathbf{R}$ and $\mathbf{t}$ are correct. At the beginning of registration, $\mathbf{R}$ and $\mathbf{t}$ are often quite inaccurate. In this case, all tentative correspondences $(\mathbf{p}_i, \mathbf{p}'_j)$ established can be considered equally unreliable, so β will be small, and all w_{ij} are close to a constant. As registration progresses, $\mathbf{R}$ and $\mathbf{t}$ become increasingly accurate, and β is asymptotically increased to penalize tentative correspondences $(\mathbf{p}_i, \mathbf{p}'_j)$ having large errors $e_{ij}^2 = ||\mathbf{p}'_j - \mathbf{R}\mathbf{p}_i - \mathbf{t}||^2$. Only those tentative correspondences $(\mathbf{p}_i, \mathbf{p}'_j)$ with small errors now have a non-zero constant weight, and those with large errors have a weight close to 0. For all the points in one dataset without correspondents in another, the resulting tentative correspondences will have large errors and thus a weight close to 0. This is a deterministic annealing process for global optimization of w_{ij}; β is called the inverse temperature.

In this objective function, the first term sums the weighted registration error $e_{ij} = ||\mathbf{p}'_j - \mathbf{R}\mathbf{p}_i - \mathbf{t}||$ for each tentative correspondence $(\mathbf{p}_i, \mathbf{p}'_j)$, while the second term is the negative entropy of these weights w_{ij} (*negative* entropy is used as this is a minimization problem). Thus, the optimization process aims to maximize the entropy (EntMax) of the weights, so that eventually, all the incorrect and correct correspondences will have the same zero and nonzero constant weights.

Differentiating the above objective function with regard to the unknown w_{ij} and setting the result to zero yields

$$w_{ij} = \exp(-\beta ||\mathbf{p}'_j - \mathbf{R}\mathbf{p}_i - \mathbf{t}||^2). \quad (2)$$

Weights and Correspondences in the Overlapping Area

Each w_{ij} in Eq. 2 describes the extent to which $(\mathbf{p}_i, \mathbf{p}'_j)$ is correct with regard to the existing rigid transformation parameters $\mathbf{R}$ and $\mathbf{t}$. However, it does not consider the situation in which in the real world, a point on the surface of an object in one view cannot correspond to more than one point in another. Let the weight a_i describe how likely it is that point $\mathbf{p}_i$ lies in the overlapping area and the point $\mathbf{p}'_{c(i)}$ be its most likely correspondent in $\mathbf{P}'$. Then both a_i and $\mathbf{p}'_{c(i)}$ need to be further refined using the basic weights estimated from Eq. 2. To this end, the simplified min-sum (SMS) algorithm [10] is used. It sends messages $m_{i\to j}$ and $m'_{j\to i}$ between points $\mathbf{p}_i$ and $\mathbf{p}'_j$ ($i = 1, 2, \cdots, n_1$; $j = 1, 2, \cdots, n_2$), so that the two-way constraint of mutual correspondence is communicated and circulated among all points: (1) if $\mathbf{p}_i$ matches $\mathbf{p}'_j$, then $\mathbf{p}'_j$ should also match $\mathbf{p}_i$; (2) if $\mathbf{p}_i$ does not match $\mathbf{p}'_j$, then $\mathbf{p}'_j$ should not match $\mathbf{p}_i$. The SMS algorithm in outline works as follows:

1. Initialize the messages $m_{i\to j}$=$m'_{j\to i}$=w_{ij}.
2. Update messages $m_{i\to j}$ from $\mathbf{P}$ to $\mathbf{P}'$ using messages $m'_{j\to i}$:
   - For each point $\mathbf{p}_i$, identify the point $\mathbf{p}'_k$ in $\mathbf{P}'$ with maximum message $m'_{k\to i}$ and also the second largest message $m'_{l\to i}$.
   - Update $m_{i\to j}$: if $j = k$, then $m_{i\to j} = w_{ij} - m'_{l\to i}$ else $m_{i\to j} = w_{ij} - m'_{k\to i}$.

3. Update messages $m'_{j\to i}$ from $\mathbf{P}'$ to $\mathbf{P}$ using messages $m_{i\to j}$:
   - For each point $\mathbf{p}'_j$, identify the point $\mathbf{p}_k$ in $\mathbf{P}$ with maximum message $m_{k\to j}$ and also the second largest message $m_{l\to j}$.
   - Update $m'_{j\to i}$: if $i = k$, then $m'_{j\to i} = w_{ij} - m_{l\to j}$ else $m'_{j\to i} = w_{ij} - m_{k\to j}$.
4. Estimate the weight a_i of point $\mathbf{p}_i$ in the overlapping area with $\mathbf{P}'$ and its most likely correspondent $\mathbf{p}'_{c(i)}$ in $\mathbf{P}'$: $a_i = \max_j m'_{j\to i}$ and $c(i) = \text{argmax}_j\, m'_{j\to i}$.
5. Repeat steps 2–4 until fewer than 1% of the points $\mathbf{p}_i$ in $\mathbf{P}$ change correspondents $\mathbf{p}'_{c(i)}$ between two successive iterations or the maximum number (taken as 5 here) of iterations has been reached.

As long as the optimal solution to a_i and $(\mathbf{p}_i, \mathbf{p}'_{c(i)})$ is unique, it is guaranteed [10] that this message-passing scheme will find it.

Estimation of the Underlying Rigid Transformation

Once the tentative correspondences $(\mathbf{p}_i, \mathbf{p}'_{c(i)})$ and their weights a_i have been estimated, the following objective function is used to reestimate the underlying rigid transformation rotation matrix $\mathbf{R}$ and translation vector $\mathbf{t}$ in a weighted least-squares sense:

$$J(\mathbf{R},\mathbf{t}) = \min_{\mathbf{R},\mathbf{t}} \sum_{i=1}^{n_1} a_i ||\mathbf{p}'_{c(i)} - \mathbf{R}\mathbf{p}_i - \mathbf{t}||^2. \quad (3)$$

This objective function can be optimized using various methods having a closed-form solution. The quaternion method [4] is one of these methods and is outlined as follows:

- Compute the weighted centroids $(\bar{\mathbf{p}}, \bar{\mathbf{p}}')$ of the tentative correspondences $(\mathbf{p}_i, \mathbf{p}'_{c(i)})$: $\bar{\mathbf{p}} = \sum_{i=1}^{n_1} a_i \mathbf{p}_i / \sum_{i=1}^{n_1} a_i$ and $\bar{\mathbf{p}}' = \sum_{i=1}^{n_1} a_i \mathbf{p}'_{c(i)} / \sum_{i=1}^{n_1} a_i$.
- Compute the weighted 3×3 covariance matrix $\mathbf{C}=\{c_{mn}\}$ of the tentative correspondences $(\mathbf{p}_i, \mathbf{p}'_{c(i)})$: $\mathbf{C} = \sum_{i=1}^{n_1} a_i (\mathbf{p}_i - \bar{\mathbf{p}})(\mathbf{p}'_{c(i)} - \bar{\mathbf{p}}')^T$.
- Compute the difference $\mathbf{D} = \{d_{mn}\}$ of matrix $\mathbf{C}$ and its transpose $\mathbf{C}^T$: $\mathbf{D} = \mathbf{C} - \mathbf{C}^T$.
- Construct a new incremental 4×4 symmetric matrix $\mathbf{A}$ using matrices $\mathbf{C}$, $\mathbf{D}$, and $\mathbf{I}$: $\mathbf{A} = \begin{pmatrix} \text{trace}(\mathbf{C}) & \Delta^T \\ \Delta & \mathbf{C} + \mathbf{C}^T - \text{trace}(\mathbf{C})\mathbf{I} \end{pmatrix}$ where $\Delta = \begin{pmatrix} d_{23} \\ d_{31} \\ d_{12} \end{pmatrix}$.
- Find the eigenvector $\mathbf{q}=(q_0\ q_1\ q_2\ q_3)^T$ of matrix $\mathbf{A}$ with the maximum eigenvalue.

$$\mathbf{R} = \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 + q_0q_2) \\ 2(q_1q_2 + q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 - q_0q_1) \\ 2(q_1q_3 - q_0q_2) & 2(q_2q_3 + q_0q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{pmatrix} \quad (4)$$

- Estimate the rotation matrix $\mathbf{R}$ from $\mathbf{q}$ using
- Estimate the translation vector $\mathbf{t}$ from $(\bar{\mathbf{p}}, \bar{\mathbf{p}}')$ and $\mathbf{R}$ as: $\mathbf{t} = \bar{\mathbf{p}}' - \mathbf{R}\bar{\mathbf{p}}$.

Summary of the Registration Algorithm

Pulling together all the ingredients described in the previous paragraphs, the registration algorithm can be summarized as follows:

Initialize $\mathbf{R}$ to the identity matrix $\mathbf{I}$, rotation axis $\mathbf{h} = (0\ 0\ 1)^T$, rotation angle $\theta = 0.1(rad)$, $\mathbf{t} = \frac{1}{n_2}\sum_{j=1}^{n_2} \mathbf{p}'_j - \frac{1}{n_1}\sum_{i=1}^{n_1} \mathbf{p}_i$, initial inverse temperature β_0, final inverse temperature β_f, the expected relative transformation estimation error ρ as defined in the Experimental Results section, maximum number of iterations, and inverse temperature $\beta = \beta_0$

Outerloop: While $(\beta_0 < \beta_f)$ do the following
Iteration number $k = 0$;
Innerloop: Do the following while the relative errors in the rotation and translation for two successive iterations are larger than a threshold ρ and while the maximum number of iterations has not been exceeded
$k \leftarrow k + 1$;
Estimate the weight w_{ij} of each tentative correspondence $(\mathbf{p}_i, \mathbf{p}'_j)$ between $\mathbf{P}$ and $\mathbf{P}'$;
Estimate the weight a_i of point $\mathbf{p}_i$ and its most likely correspondent $\mathbf{p}'_{c(i)}$ in the overlapping area;
Estimate the rigid transformation parameters, the rotation matrix $\mathbf{R}$ and translation vector $\mathbf{t}$

Endinnerloop

$\beta \leftarrow \beta_r \beta$

Endouterloop

If better estimates are available for $\mathbf{R}$ (e.g., from a turntable) and $\mathbf{t}$, they should be used for initialization. In the experiments to be described later, the following values were used: $\beta_0 = n_1 n_2 / \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} ||\mathbf{p}'_j - \mathbf{p}_i||^2$, $\beta_r = 1.06$, $\beta_f = 4{,}000\beta_0$, and $\rho = 0.001$, and ten iterations were used at most. These values may be data dependent and thus, need to be fine-tuned for better results, even though the dependence may be weak.

The above algorithm is called the message-passing two-way (MPTW) constraint algorithm. From its development, it can be seen that it has a computational complexity of $O(n_1 n_2)$ for estimating w_{ij}, $O(n_1 n_2)$ for message passing, and $O(n_1)$ for estimating the underlying transformation. Thus, it has an overall complexity of $O(n_1 n_2)$. Experimental results produced by this method are given later in this chapter.

Application

Rigid registration finds numerous applications in digital shape reconstruction, anatomical modeling, robot navigation, object recognition, and inspection, to just name a few. These applications can be classified into three main categories according to the information sought:

- Merging information from different views for shape modeling. Rigid registration is applied in [11] to data captured by a flying laser range scanner from different viewpoints. Registration allows redundant points in overlapping areas to be removed and the remaining points to be integrated, leading to a full digital 3D model. This has been used to create a digital shape archive documenting the Bayon temple. In an example of human anatomical modeling, data from computerized tomography, video cameras, and a controllable XYZ coordinate table are registered in [12] using Image-Pro Plus software (Media-Cybernetics, Inc., www.mediacy.com) to allow modeling of the entire masticatory system.
- Using underlying transformations for localization. Different range scans from an AIS laser range finder are aligned in [13] to allow localization of a robot and avoidance of obstacles during the process of navigation. To provide the requisite accuracy, the data are first aligned using a fast variant of the traditional iterative closest point (ICP) algorithm [4], then multiple scans are registered simultaneously for global consistency, and finally dynamic objects are detected by using additional sensors together with the information from registered frames.
- Making comparisons between an original view or CAD model and a new view for recognition and inspection. Different range images are registered in [14] that allow an ear detection rate of 99.9% and an ear identification rate of 95.4% on Collection F of the UND database. To perform recognition, ear images are detected from 2D profile images using a cascaded AdaBoost detector. The corresponding 3D ear data is then extracted from the co-registered range image and represented with local 3D features. Local features are used to construct a rejection classifier, to extract a minimal region with feature-rich data points, and finally, to compute the initial transformation for matching with the ICP algorithm. Five laser scans acquired at different stages of the construction of the steel structure of a building are registered against its 3D CAD model containing 612 objects with a total of 19,478 facets in [15] for object recognition and calculation of objects' as-built dimensions and dimensional compliance control. The ICP algorithm is adapted for the global registration. A CAD object is recognized when the surface covered by its registered as-built point cloud is larger than a threshold. For better pose estimation, the ICP variant is applied again between the matched data and model points individually for each object. Once the as-built poses of all recognized objects are calculated, they are compared to their as-designed poses in order to infer some information on the compliance of the project with respect to dimensional tolerances.

Open Problems

Automatic registration is challenging, and the following issues remain open:

- Point sampling. When a scanner is used to image an object of interest, it essentially samples the object surface. In this case, the points sampled in different views of the surface are sampled at different locations. The nature of this imaging process implies

that no exact point will be sampled from two different viewpoints. The variation in location of the sampled points is determined by the resolution of the scanner. The higher the resolution, the smaller the variation. Even though the sampling resolution can be increased, it has a limit. Consequently, no matter how good the scanner is, typically, correspondences are never correct and must have some errors.

– Robustness. Outliers are a big problem in scan data. They come from either background, other parts of the same surface, or other objects, or are artiefacts of the scanning method. In practice, numbers of outliers can vary, and they can be clustered rather than spread uniformly through the data. Robust algorithms need to take such issues into account. The unpredictable proportion, distribution, and location of outliers render it difficult to evaluate the relative quality of the established correspondences. To this end, a large number of algorithms have been developed based on feature extraction and matching [2], the optimized match of points in shapes [3, 4, 6], and randomized underlying transformation search [5]. All these methods can succeed in some cases but may fail catastrophically in others. These failures have two main causes: (1) these algorithms all depend upon some parameters, many of which are data dependent, and (2) no prior knowledge is available about such things as distribution of points, occlusion, appearance and disappearance of points, imaging noise, and the magnitude of the transformation. If such knowledge is available, expected locations of overlapping and nonoverlapping points can be used to guide the registration process.
– Computational efficiency. In theory, only three correct exact correspondences are needed for the estimation of the underlying transformation. In practice, since the correspondences are usually contaminated by noise, more correspondences should be used instead to obtain a consensus. On the other hand, the more measured points there are for each shape to be registered, the more candidates that must be considered in the search for correct correspondences. Various algorithms exist for searching the space of tentative correspondences, with various computational complexities. Both the max-product algorithm [16] and the game-theoretic approach [17] have a computational complexity of $O(n^4)$, while the simplified min-sum algorithm [10] and the brute force iterative closest point (ICP) algorithm [4] have a computational complexity of $O(n^2)$, and the variant of the ICP algorithm in [6] has a computational complexity of $O(n \log n)$. These algorithms all have greater than linear time complexity. Both the resolution and field of view of scanners have increased steadily in the last decade, and now millions of points can be captured within minutes. While such a large number of points are useful to accurately capture the geometry and fine details of shapes, it causes difficulties for registration. An algorithm with a computational complexity of $O(n^2)$ can be practically applied interactively to shapes with thousands of points, and one with a computational complexity of $O(n \log n)$ can be used with tens of thousands of points. Other cases take tens of minutes upwards to perform registration and thus cannot satisfy the requirements of real-time applications.
– Performance measurement. Various methods have been developed in the literature to assess the performance of registration algorithms: the root-mean-squared-distance (RMSD) [7, 17], mean squared error (MSE) [5, 15], the average of registration errors of reciprocal correspondences [6], or the surface interpenetration measure (SIM) [5]. However, no single parameter can always successfully capture the performance of different algorithms with different data. All these methods measure different aspects of the success of registration algorithms. A further problem arises when the underlying rigid transformation is modeled as nonrigid [18], using, e.g., thin-plate splines (TPS). In such cases, these performance measures may indicate low errors, but this does not necessarily mean that the registration is accurate. This can happen because the TPS has $n + 4$ degrees of freedom and often overfits any noisy points and outliers in the data.

Experimental Results

In this section, both synthetic data and a pair of overlapping real measurements of a toy cow (cow45, cow48) [1] are used to validate the MPTW algorithm given earlier. In this case, ground truth values of the transformation parameters are known.

Synthetic Shape Generation

The synthetic data were generated as follows. Firstly, n points $\mathbf{P} = \{\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_n\}$ were randomly generated with uniform distribution within the 3D space $[5, 15] \times [5, 15] \times [5, 15]$. These points were then subjected to a rotation angle θ around a fixed unit rotation axis $\mathbf{h} = (h_x \ h_y \ h_z)^T$ randomly placed and oriented with uniform distribution within the 3D space $[1, 3] \times [1, 3] \times [1, 3]$ followed by a constant translation vector $\mathbf{t}$ randomly generated with uniform distribution within the 3D space $[10, 20] \times [10, 20] \times [10, 20]$. Let the transformed points be $\mathbf{P}' = \{\mathbf{p}'_1, \mathbf{p}'_2, \cdots, \mathbf{p}'_n\}$, where $\mathbf{p}'_i = \mathbf{R}\mathbf{p}_i + \mathbf{t}$, $\mathbf{R} = \mathbf{I} + \mathbf{H}\sin\theta + \mathbf{H}^2(1 - \cos\theta)$, and $\mathbf{H} = \begin{pmatrix} 0 & -h_z & h_y \\ h_z & 0 & -h_x \\ -h_y & h_x & 0 \end{pmatrix}$. Given this data, precise knowledge of the selected points and their correspondents $(\mathbf{p}_i, \mathbf{p}'_i)(i = 1, 2, \cdots, n)$ is available. The underlying transformation given by $\mathbf{R}$ and $\mathbf{t}$ serves as ground truth for error estimation and validation of the algorithm.

In order to simulate noisy real-world data, Gaussian white noise was added to the coordinates of each point with standard deviation $\sigma_1 = 0.1$ in one series of experiments and $\sigma_2 = 0.2$ in another. In order to simulate occlusion, appearance, and disappearance of points in both $\mathbf{P}$ and $\mathbf{P}'$, the last 40% points in $\mathbf{P}$ and the first 30% points in $\mathbf{P}'$ were removed, giving two new sets of points $\mathbf{P}$ and $\mathbf{P}'$ for registration with 30% overlap in 3D. (Note that this simulates a new scenario where appearing and disappearing points can appear anywhere in 3D space, instead of at only boundaries of shapes. Such different scenarios are useful to test the robustness of registration algorithms.)

Performance Measurement

The performance of the MPTW algorithm was measured in three ways:

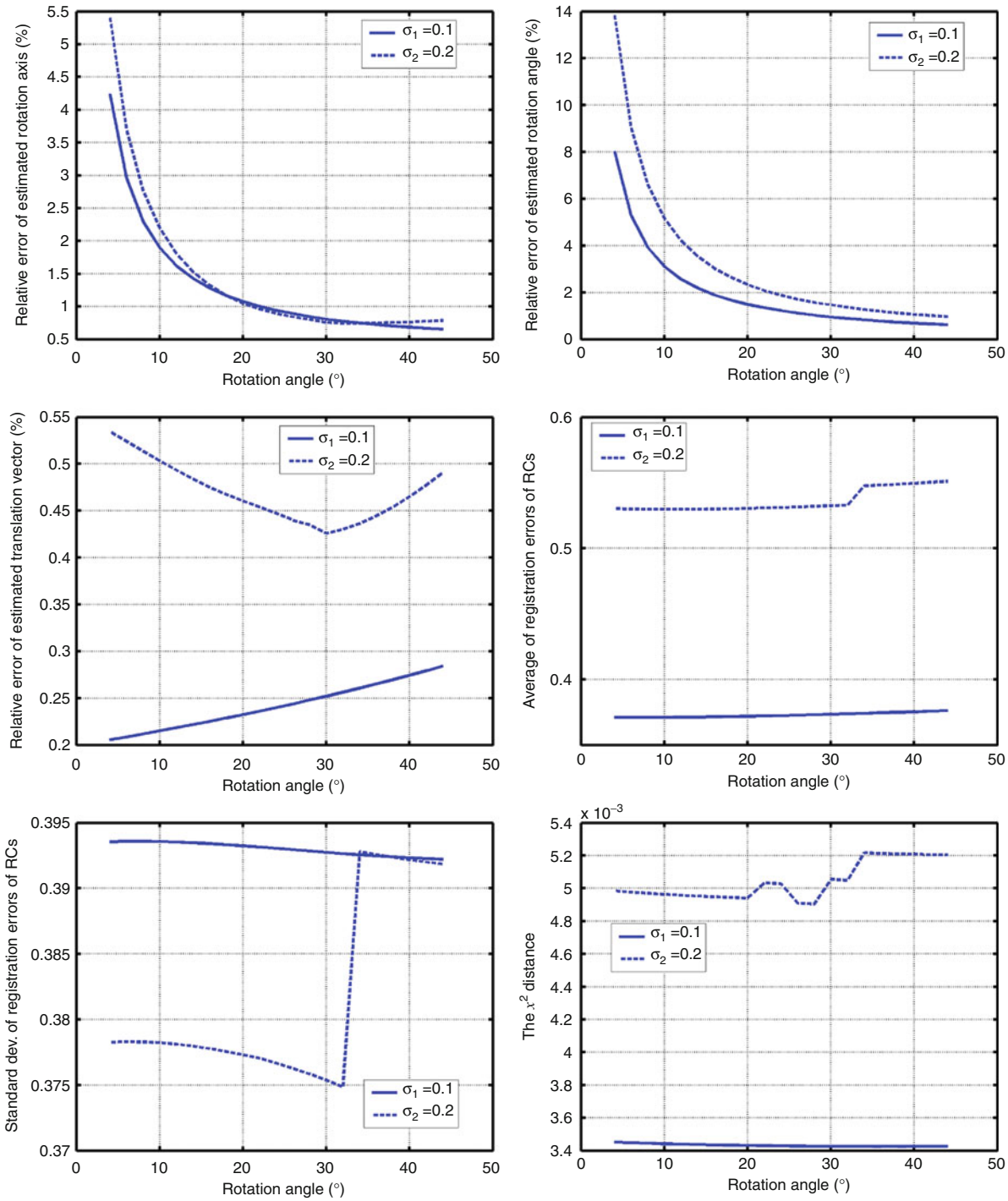
- The percentage relative errors $e_{\mathbf{h}}$, e_θ, and $e_{\mathbf{t}}$ in estimated rotation axis $\hat{\mathbf{h}}$, rotation angle $\hat{\theta}$, and translation vector $\hat{\mathbf{t}}$ for the underlying transformation: $e_{\mathbf{h}} = ||\hat{\mathbf{h}} - \mathbf{h}|| \times 100\%$, $e_\theta = (\hat{\theta} - \theta)/\theta \times 100\%$, and $e_{\mathbf{t}} = ||\hat{\mathbf{t}} - \mathbf{t}||/||\mathbf{t}|| \times 100\%$, where $\hat{\theta} = \cos^{-1}((r_{11} + r_{22} + r_{33} - 1)/2)$, $\hat{\mathbf{h}} = (r_{32} - r_{23}, r_{13} - r_{31}, r_{21} - r_{12})^T / \sin\hat{\theta}$, and $\hat{\mathbf{R}}$ and $\hat{\mathbf{t}}$ are estimated rotation matrix and translation vector. The relative error e_r of the rotation vector is defined as $e_r = ||\hat{\theta}\hat{\mathbf{h}} - \theta\mathbf{h}||/\theta \times 100\%$.
- The average e_μ and standard deviation e_σ of registration errors of the finally established reciprocal correspondences (RCs) and the χ^2 distance d: $e_\mu = \frac{1}{N}\sum_{i=1}^{N} e_i$, $e_\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(e_i - e_\mu)^2}$, and $d = \frac{1}{N^2}\sum_{j=1}^{N}\sum_{i=1}^{N}(c_{ij} - 1)^2/(c_{ij} + 1)$, where N is the number of RCs, $e_i = ||\mathbf{p}'_{c(i)} - \hat{\mathbf{R}}\mathbf{p}_i - \hat{\mathbf{t}}||$, $c_{ij} = \min(||\mathbf{p}_j - \mathbf{p}_i||, ||\mathbf{p}'_{c(j)} - \mathbf{p}'_{c(i)}||)/\max(||\mathbf{p}_j - \mathbf{p}_i||, ||\mathbf{p}'_{c(j)} - \mathbf{p}'_{c(i)}||)$, and $(\mathbf{p}_i, \mathbf{p}'_{c(i)})$ and $(\mathbf{p}_j, \mathbf{p}'_{c(j)})$ are any two established RCs. The χ^2 distance measures the extent to which the established RCs are compatible in the sense of rigidity: $||\mathbf{p}_j - \mathbf{p}_i|| = ||\mathbf{p}'_{c(j)} - \mathbf{p}'_{c(i)}||$. An RC is a correspondence $(\mathbf{p}_i, \mathbf{p}'_{c(i)})$ that if point $\mathbf{p}_i$ in $\mathbf{P}$ corresponds to $\mathbf{p}'_{c(i)}$ in $\mathbf{P}'$, then $\mathbf{p}'_{c(i)}$ in $\mathbf{P}'$ also corresponds to $\mathbf{p}_i$ in $\mathbf{P}$. Clearly, RCs satisfy the two-way constraint and are thus (more) likely to represent correct correspondences.
- Computational time. All the experiments were carried out on a Pentium IV 2.80 GHz computer with 512 M memory and implemented using the C programming language inside Microsoft Visual C++ 6.0.

While the first kind of measurements requires ground truth, the second does not. The smaller these values are, the more accurate the registration algorithm is.

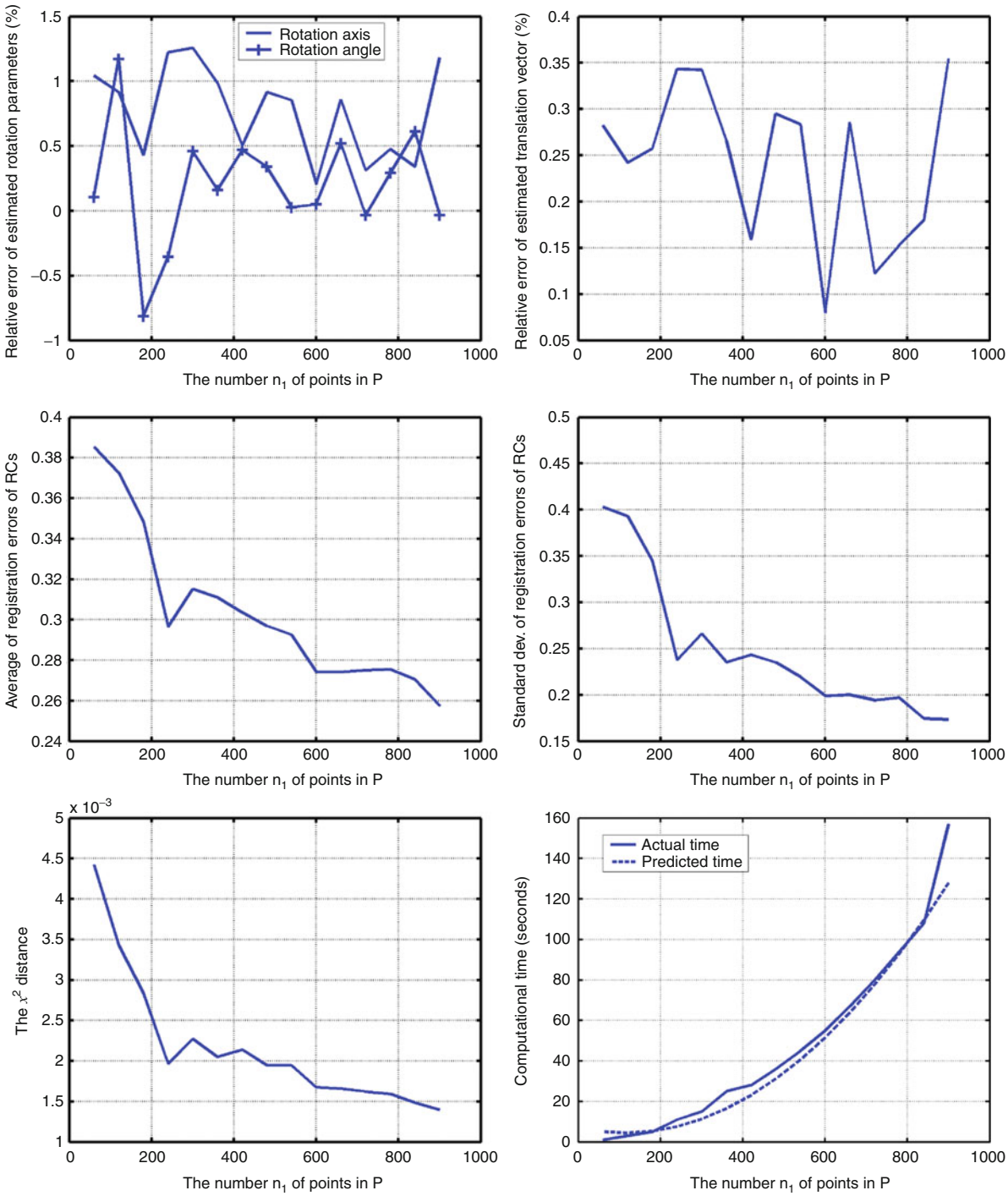
Results and Analysis

The experimental results for synthetic data are presented in Figs. 3 and 4 and Tables 1 and 2. Solid lines correspond to a lower level $\sigma_1 = 0.1$ of noise, while dashed lines correspond to a higher level $\sigma_2 = 0.2$ of noise. In Fig. 3 and Table 1, n was fixed as $n = 200$ with the rotation angle θ of the underlying transformation ranging from 4° to 44° at intervals of 2°. In Fig. 4 and Table 2, the rotation angle θ of the underlying transformation was fixed as $\theta = 25°$ with the parameter n ranging from 100 to 1,500 at intervals of 100, and the data were subject to the lower level $\sigma_1 = 0.1$ of noise.

From Fig. 3 and Table 1, it can be seen that all the parameters of interest smoothly vary with the rotation angle of the underlying transformation, showing that the MPTW algorithm is very stable. When data were corrupted by stronger noise, the MPTW algorithm typically produced worse registration results, as might be expected. The estimation of the translation vector is usually more accurate than that of either the rotation axis or angle. The reason why the χ^2 distance is small



Rigid Registration, Fig. 3 Different performance parameters as a function of the rotation angle of the underlying transformation. *Top left*: rotation axis; *Top right*: rotation angle; *Middle left*: translation vector; *Middle right*: the average of the registration errors of RCs; *Bottom left*: standard deviation of registration errors of RCs; *Bottom right*: the χ^2 distance



Rigid Registration, Fig. 4 Different performance measures as a function of the number n_1 of points in the first shape **P**. *Top left*, rotation parameters; *top right*, translation vector; *middle left*, average of registration errors of RCs; *middle right*, standard deviation of registration errors of RCs; *bottom left*, the χ^2 distance; *bottom right*, computational time

Rigid Registration, Table 1 Average μ and standard deviation σ of the relative estimation errors $e_{\mathbf{h}}$, e_θ, and $e_{\mathbf{t}}$ of the rotation axis $\hat{\mathbf{h}}$, rotation angle $\hat{\theta}$, and translation vector $\hat{\mathbf{t}}$, average e_μ and standard deviation e_σ of registration errors of RCs, and the χ^2 distance of the MPTW algorithm as a function of the rotation angle θ of the underlying transformation

Noise	Meas	$e_{\mathbf{h}}(\%)$	$e_\theta(\%)$	$e_{\mathbf{t}}(\%)$	e_μ	e_σ	χ^2 Dis
0.1	μ	1.30	1.95	0.24	0.37	0.39	0.0034
	σ	0.88	1.80	0.02	0.00	0.00	0.0000
0.2	μ	1.45	3.20	0.47	0.54	0.38	0.0050
	σ	1.17	3.14	0.03	0.01	0.01	0.0001

Rigid Registration, Table 2 Average μ and standard deviation σ of relative estimation errors $e_{\mathbf{h}}$, e_θ, and $e_{\mathbf{t}}$ of rotation axis $\hat{\mathbf{h}}$, rotation angle $\hat{\theta}$, and translation vector $\hat{\mathbf{t}}$; average e_μ and standard deviation e_σ of registration errors of RCs; the χ^2 distance and the computational time in seconds of the MPTW algorithm as a function of the number n_1 of points in the first shape $\mathbf{P}$

Meas	$e_{\mathbf{h}}(\%)$	$e_\theta(\%)$	$e_{\mathbf{t}}(\%)$	e_μ	e_σ	χ^2 Distance	time (s)
μ	0.77	0.20	0.24	0.30	0.25	0.0022	48.67
σ	0.34	0.44	0.08	0.04	0.07	0.0008	43.70

is that the compatibility of RCs increases as a function of their Euclidean distances when they are far apart, even though they may have the same registration error.

From Fig. 4 and Table 2, it can be seen that all the parameters of interest fluctuate somewhat as a function of the number n_1 of points in $\mathbf{P}$. This is because the points were randomly generated each time and thus their distribution changed accordingly. The overall trend is clear and stable. The actual computational time exhibits quadratic behaviour in the number n_1 of points in $\mathbf{P}$, fitting a second order polynomial: $time = 0.0002n_1^2 - 0.0458n_1 + 7.0813$ s. Note that in this case, both $\mathbf{P}$ and $\mathbf{P}'$ have similar numbers of points. If n_1 were 10,000, it would have taken 19,549 s to register the shapes: despite its accuracy and robustness, the MPTW algorithm is unsuited to the registration of shapes with a large number of points. To make it practically useful, either the numbers n_1 and n_2 of points from $\mathbf{P}$ and $\mathbf{P}'$ must be reduced by sampling or fewer combinations of these points must be considered for initial weight estimation and message passing.

The cow45 and cow48 shapes have a transformation with an expected rotation angle of 30° around an unknown rotation axis and include 7,049 and 3,195 points with an average of the distances between neighboring points being 1.66 mm and 1.73 mm, respectively. The estimated rotation matrix $\hat{\mathbf{R}}$ and translation vector $\hat{\mathbf{t}}$ are $\hat{\mathbf{R}} = \begin{pmatrix} 0.86 & -0.43 & 0.29 \\ 0.42 & 0.90 & 0.09 \\ -0.30 & 0.05 & 0.95 \end{pmatrix}$ and $\hat{\mathbf{t}} = (389.74\ 115.65\ -61.49)^T$, from which the rotation axis $\hat{\mathbf{h}}$ and angle $\hat{\theta}$ are $\hat{\mathbf{h}} = (-0.04\ 0.57\ 0.82)^T$ and 31.11°, respectively. The relative error in the estimated rotation angle is 3.70%. The number N of RCs found is $N = 1643$, while the average and standard deviation error in the RCs are 0.80 and 0.64 mm. This is a sub-pixel accuracy that a registration algorithm can at best achieve. The χ^2 distance is 0.0006. The superimposition of the transformed cow45 and cow48 shapes is presented in Fig. 1 – the yellow color represents the transformed cow45 shape $\mathbf{P}$, and the green color represents the reference cow48 shape $\mathbf{P}'$. Even though no a priori knowledge was given about what the two point clouds represent, and they have an overlap of just $1{,}643/\max(7{,}049, 3{,}195)$ points = 23.31%, they have been brought into a good alignment with a large amount of interpenetration. The disappearing cluttered background and the appearing left front leg have been correctly identified. The computational time taken for registration was 7,595 s, which is less than predicted: 9,622 s, resulting in a prediction error of 21%. This is because the relative difference in the numbers of points in $\mathbf{P}$ and $\mathbf{P}'$ is larger in the real data than in the synthetic data.

References

1. OSU(MSU/WSU) range image database (2010) http://sampl.ece.ohio-state.edu/data/3DDB/RID/index.htm
2. Johnson A, Hebert M (1999) Using spin images for efficient object recognition in cluttered 3D scenes. IEEE Trans Pattern Anal Mach Intell 21:433–449
3. Gold S, Rangarajan A et al (1998) New algorithms for 2-D and 3-D point matching: pose estimation and correspondence. Pattern Recognit 31:1019–1031
4. Besl PJ, McKay ND (1992) A method for registration of 3-D shapes. IEEE Trans Pattern Anal Mach Intell 14(2): 239–256

5. Silva L, Bellon Olga RP, Boyer KL (2005) Precision range image registration using a robust surface interpenetration measure and enhanced genetic algorithms. IEEE Trans Pattern Anal Mach Intell 27:762–776
6. Liu Y (2005) Automatic 3d free form shape matching using the graduated assignment algorithm. Pattern Recognit 38:1615–1631
7. Phillips JM, Liu R, Tomasi C (2007) Outlier robust ICP for minimizing fractional RMSD. In: Proceedings of International Conference on 3D Digital Imaging and Modeling. IEEE Computer Society, Washington DC, pp 427–434
8. Chen Y, Medioni G (1992) Object modelling by registration of multiple range images. Image and Vision Computing, 10(3):145–155
9. Rusinkiewicz S, Levoy M (2001) E_cient variants of the ICP algorithm proceedings of 3rd international conference on 3-D digital imaging and modeling, pp 145–152
10. Bayati M, Shah D, Sharma M (2008) Max-product for maximum weight matching: convergence, correctness, and LP duality. IEEE Trans Inf Theory 54(3):1241–1251
11. Banno A, Masuda T, Oishi T, Ikeuchi K (2008) Flying laser range sensor for large-scale site-modeling and its applications in Bayon digital archival project. Int J Comput Vis 78:207–222
12. Pileicikien G, Surna A et al (2009) A new technique for the creation of a higher accuracy 3D geometrical model of the human masticatory system. Mechanika 78(4): 44–50
13. Surmann H, Nuchter A, Hertzberg J (2003) An autonomous mobile robot with a 3D laser range finder for 3D exploration and digitalization of indoor environments. Robot Auton Syst 45:181–198
14. Islam SMB, Davies R, Bennamoun M, Mian AS (2011) Efficient detection and recognition of 3D ears. Int J Comput Vis 95:52–73
15. Bosch F (2010) Automated recognition of 3D CAD model objects in laser scans and calculation of as-built dimensions for dimensional compliance control in construction. J Adv Eng Inform 24(1):107–118
16. Xiao P, Barnes N, Lieby P, Caetano T (2009) Applying sum and max product algorithms of belief propagation to 3D shape matching and registration. In: Proceedings of International Conference on Digital Image Computing: Techniques and Applications. IEEE Computer Society, Washington DC, pp 387–394
17. Albarelli A, Rodola E, Torsello A (2010) A game-theoretic approach to fine surface registration without initial motion estimation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, pp 430–437
18. Chui H, Rangarajan A (2003) A new point matching algorithm for non-rigid registration. Comput Vis Image Underst 89:114–141

Rigid Transformation Estimation

▸Rigid Registration

Robot-Camera Calibration

▸Hand-Eye Calibration

Robust Clustering

▸Robust Methods

Robust Methods

Peter Meer[1] and Sushil Mittal[2]
[1]Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ, USA
[2]Department of Statistics, Columbia University, New York, NY, USA

Synonyms

Robust clustering; Robust regression

Definition

The goal of robust methods in computer vision is to extract all the information necessary to solve a given task while discarding everything that is not needed. The tasks can be very simple or very complex, but in real-life applications, a robust procedure will always be required. In the end, the performance of a machine for solving a vision problem will be judged against that of human observers performing the equivalent task. Since the human visual system works in a much more sophisticated manner than the present day computer vision systems, this ultimate goal is still far from reality.

Background

The task of a robust algorithm is to derive a *model* that estimates one or more structures present in the data, each structure depending only on a part of the data. A point obeying the model is called an *inlier*, while a point not obeying it is called an *outlier*. Since

there are several inlier structures, therefore relative to one inlier structure the outliers can either be on another inlier structure (structured outliers) or can be completely unstructured. The model estimation problems can be divided into two categories – location estimation (clustering) and regression estimation.

Nonparametric method-based image segmentation is a typical example of a location estimation problem. In this class of problems, first the data points in the original space are transformed to a feature space (suitable for the particular problem), where they form *meaningful* clusters. The goal of a robust algorithm then is to find correct estimates of the cluster centers and associate observed data points to their respective clusters while rejecting the outliers. Among nonparametric clustering algorithms, kernel density estimation-based methods are used very frequently for vision applications. These methods start by computing a kernel density estimate over the entire observed data. The modes of this density correspond to the cluster centers of the data in the feature space. These modes, together with their associated points, are then recovered using techniques such as mean shift [1]. The bandwidth of the kernel is a crucial parameter which is either supplied by the user or determined from the data prior to estimating the density.

Robust regression is the second class of estimation problems. Currently, in computer vision, a maximum of four or five regression inlier structures can be estimated from the data. The structures are usually recovered iteratively, one after the other. Almost all the models in geometric vision describing 3D relations, 3D to 2D relations, or 2D to 2D relations are nonlinear. For example, the epipolar geometry between corresponding points in two frames, that is, the fundamental or essential matrix estimation, has to satisfy the constraint $\mathbf{y}_2^T \mathbf{F} \mathbf{y}_1 = 0$. Given a few point correspondences between the two frames, a 3×3 matrix $\mathbf{F}$ has to be determined. The equation written in affine coordinates has four variables – the x, y coordinates of the matched point in the two frames. However, due to errors in the output of a point matching algorithm, the points may or may not be in correspondence.

In most of the cases, a robust computer vision estimator starts by reinterpreting the nonlinear relation (generally) in a higher-dimensional linear space. The original variables and their monomials which appear in products become the linearized variables $\mathbf{x}_j$. In the case of fundamental matrix estimation, the derived linear equation has eight variables, the four original variables and the four monomials – the products terms formed by multiplying each coordinate from the first image with each coordinate from the second image. See *Epipolar Constraint* entry.

The linearized variables are much easier to estimate, and the inliers can be obtained using a robust procedure. However, the original problem almost certainly has additional constraints which also have to be satisfied. A nonlinear processing, taking into account all the constraints, leads to the final estimate. For data containing multiple inlier structures, the whole procedure is repeated for each structure until all of the structures have been recovered. A good robust algorithm should be able to reliably extract all the inlier structures (in general having different inlier noise) *without* any prior knowledge about the actual number of structures present.

The epipolar geometry example mentioned above has a simple linearized constraint. But more complex cases also exist. In camera calibration (or camera pose recovery), there are five original variables – the coordinates of a point in 3D and the corresponding coordinates in 2D. After linearization, the constraint becomes a 2×12 matrix for each point pair. In projective factorization over F frames, with all points visible in each frame, the original and the linearized variables have the same dimensions – a $3F \times (3F - 3)$ matrix. See the respective entries.

Due to linearization, the first-order approximation of the covariance matrix of the linearized variables is heteroscedastic, that is, each of the points has a different covariance even when the original variables are homoscedastic, that is, have the same variance. The heteroscedasticy of the linearized variables is not taken into account in majority of the robust algorithms. Robust estimators used in computer vision should be unbiased in the first-order approximation. This means that an estimate should converge toward its theoretical value (up to first order) when the average in taken over a large number of inliers. Heteroscedasticy plays an important role in achieving that.

The following procedure shows how to estimate each inlier structure. After the linearization, let the subscript "o" denote the theoretical value of the m-dimensional linearized variable $\mathbf{x}_{io}$. In the most general case, an $m \times k$ matrix, $\boldsymbol{\Theta}$, and a k-dimensional intercept, $\boldsymbol{\alpha}$, have to be estimated. Their theoretical values are obtained using

$$\boldsymbol{\Theta}^T \mathbf{x}_{io} - \boldsymbol{\alpha} = \mathbf{0}_k \qquad i = 1, \ldots, n \,. \tag{1}$$

If Eq. (1) is solved correctly, $n_1 < n$ inliers are returned together with the estimates $\hat{\boldsymbol{\Theta}}$ and $\hat{\boldsymbol{\alpha}}$ and the corrected points $\hat{\mathbf{x}}_i$. If required, the covariances of these points can also be computed.

Many times, the constraint $\boldsymbol{\Theta}^T \boldsymbol{\Theta} = \hat{\boldsymbol{\Theta}}^T \hat{\boldsymbol{\Theta}} = \mathbf{I}_{k \times k}$ is also used during the estimation procedure since it reduces the ambiguity of the estimate $\hat{\boldsymbol{\Theta}}$ only to an orthonormal gauge matrix [2]. This constraint defines a k-dimensional subpace of the n-dimensional space and thus lies on a Grassmann manifold. While exploiting group theoretic methods for processing over Grassmann manifolds is complicated, they can be used to estimate a better local optimum [15]. A nonlinear variant of mean shift has also been used to perform clustering over Grassmann manifolds [3].

Important Methods

Robust computer vision and robust statistics appeared independently at the end of 1950s and beginning of 1960s. In computer vision, the Hough transform was introduced in 1959 in a conference, and it was patented in United States in 1962. See *Hough Transform* entry.

While the rejection of outliers was practiced for hundreds of years in statistics, the first so-called robust estimator, the M-estimator, was proposed by Huber only in 1964 [4]. There are many different types of M-estimators, but each of them requires an estimate of the scale (variance) of the inlier noise beforehand. In regression problems, the estimation of scale is equivalent to the kernel bandwidth estimation in clustering. Today, in computer vision, only few M-estimator types are still used.

The least median of squares (LMedS) was introduced by Rousseeuw in 1984 and was in fashion in computer vision too in the 1990s. Today, LMedS is no longer used. While the scale of the inlier noise estimate is not needed beforehand, the LMedS can recover only a single inlier structure containing more than 50 % of the total points present in the data. The statistics of LMedS is described in the book [5], while the paper [6] also presents some of its computer vision applications. A survey of M-estimators and LMedS, including their limitations, can be found in [7], while the special issue of Computer Vision and Image Understanding [8] gives a good overview of the robust procedures of the 1990s.

In computer vision, the M-estimators in general are employed with a nonconvex, redescending loss function, which, at least theoretically, does not always converge to the global optimum solution. The prototype of this loss function class is

$$0 \le \rho(u) \le 1 \quad |u| \le 1 \qquad\qquad \rho(u) = 1 \quad |u| > 1 \,. \tag{2}$$

It is nonnegative with $\rho(0) = 0$, even symmetric with $\rho(-u) = \rho(u)$, and nondecreasing with $|u|$. Examples of several loss functions can be found in [4, 7].

In the simplest case, given the measured points and their the m-dimensional linearized values $\mathbf{x}_i$, an m-dimensional vector $\boldsymbol{\theta}$ and a scalar intercept α need to be estimated. The theoretical formulation is

$$\underset{\alpha, \boldsymbol{\theta}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \rho\left(\frac{\boldsymbol{\theta}^T \mathbf{x}_{io} - \alpha}{s} \right), \tag{3}$$

where the user has to supply *beforehand*, the scale s of the inlier noise. The residual of $\mathbf{x}_i$ is usually nonzero and is used to classify it either as an inlier or an outlier, depending on the value of scale and the loss function. Since $\rho(u)$ is symmetrical in the residuals, loss function in u^2 can be introduced instead. More complex M-estimators can then be represented with a squared Mahalanobis distance as the argument of $\rho(u^2)$. The $k \times k$ scale matrix $\mathbf{S}$ is diagonal and like before has to be supplied by the user.

The estimation is performed iterative. The initial parameter estimate is obtained by ordinary or total least squares. The nonnegative weights

$$w(u_i) = \frac{1}{u_i} \frac{d\rho(u_i)}{du} \ge 0 \qquad u_i = \frac{\boldsymbol{\theta}^T \mathbf{x}_{io} - \alpha}{s} \qquad i = 1, \ldots, n \tag{4}$$

are then computed. The values of these weights decrease from the maximum at $u = 0$ to zero at $u = \pm 1$. Beyond $u = \pm 1$, the weights are zero, and the corresponding points are considered outliers. A $n \times n$ diagonal matrix of the weights is defined, and the updated model estimate is computed taking the weights for all points into consideration. In practice, the procedure converges after a few iterations.

For any M-estimator, at least three disadvantages can be enumerated. Firstly, often it is difficult to give a good estimate of the scale of the inlier noise. For example, when the estimation is performed over a video sequence, the real scale may change, and it is very difficult to guess that change online. Secondly, in every iteration, *all* the samples participate in the process, and if the initial estimate is completely wrong, then it is very hard to recover. Thirdly, the majority of M-estimators reject only a few outliers and do not work with asymmetric noise or with data having multiple inlier structures. Figure 1 shows a line fitting example using M-estimator. While the estimate is good in case of symmetric outliers around the inliers, it is grossly incorrect when the outliers are little asymmetric.

Today in computer vision, almost everybody uses the *RANdom SAmple Consensus* (RANSAC), proposed by Fischler and Bolles in 1980 and published in a journal in 1981 [9]. RANSAC started to take over around 1995, and in 2006, a workshop to honor 25 years of the initial publication was held in conjunction with the conference on Computer Vision and Pattern Recognition.

Like M-estimator, the scale of inlier noise is still provided by the user *beforehand* in RANSAC, but the estimation is based on a totally different principle. Unlike M-estimator, in RANSAC, a hypothesis is drawn exclusively from an *elemental subset*. An elemental subset is the smallest number of data points required to fully estimate a model. The model being linear Eq. (1), the number of points in the elemental subset should be equal to the number of unknowns in the linearized model equation. For example, in the scalar case of the fundamental matrix, a hypothesis needs eight points because the 3×3 matrix $\mathbf{F}$, written as a vector, has eight independent linear unknowns. See *8-point algorithm* entry. Other hypotheses also exist, like the 7-point algorithm for fundamental matrices which takes into account the rank-two constraint of $\mathbf{F}$ too. The method requires the satisfaction of additional, in general, nonlinear constraints leading to more complicated computations. Using the same notation as defined in Eq. (3), the simplest theoretical objective function is

$$\underset{\alpha,\boldsymbol{\theta}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \rho_{zo}\left(\frac{\boldsymbol{\theta}^T \mathbf{x}_{io} - \alpha}{s}\right). \tag{5}$$

The value of ρ_{zo}, loss function for RANSAC is

$$\rho_{zo}(u) = 0 \quad |u| \leq 1 \qquad\qquad \rho_{zo}(u) = 1 \quad |u| > 1, \tag{6}$$

and it is relative easy to tune RANSAC to detect an inlier structure with constant scale.

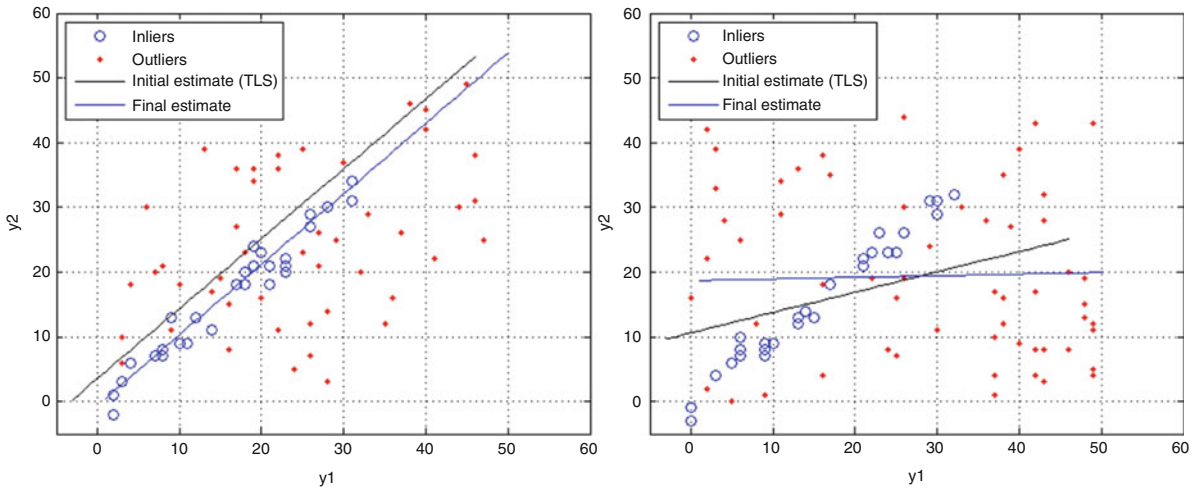
Before the start of the procedure, two parameters have to be provided to RANSAC. One is crucial – the scale of the inlier noise s, and one is approximate – the number of trials N. The value N should be several times the minimum number of trials computed for the lowest bound of the inlier/outlier ratio assumed in the problem. The number of trials increases quasi-exponentially with the size of the elemental subset. Depending on a specific problem, the number of trials can vary from a few hundreds to a few thousands.

The procedure is very simple.

- Repeat N times:
  - Choose an elemental subset by random sampling *without* replacement.
  - Find analytically, the corresponding model candidate pair $\boldsymbol{\Theta}, \boldsymbol{\alpha}$.
  - Assuming that this model candidate is valid for the entire dataset of n points, compute the value of the robust objective function. The scale estimate s is used to separate the inliers from the outliers.
- The model candidate yielding the smallest objective function value gives the estimate of RANSAC, the values $\hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\alpha}}$.

Since the constraints in the *original* problem usually impose additional limits, the detected inliers are further processed with a generally nonlinear procedure. RANSAC is a universal flowchart valid for any differentiable or nondifferentiable objective function. For example, the non-robust least squares can also be computed this way but without closed-form analytical results.

Like any elemental subset-based estimation method, RANSAC does not guarantee to find the global optimum. If successful, RANSAC gives a good inlier/outlier separation which almost always is sufficient. This separation also depends on the complexity of the objective function. For example, the scalar objective function of the epipolar geometry may not be able to eliminate all the outliers from the region detected as inliers. Additionally it could fail when



Robust Methods, Fig. 1 Line fitting example using M-estimator. *Left*: the method works in case of symmetrical outliers. *Right*: it fails in the presence of unsymmetric outliers

the data contains a lot of outliers, asymmetric noise, and/or multiple inlier structures. For example, Fig. 2 shows a simple line fitting example using RANSAC. Note that in both cases, the true value of the scale was provided. Also, traditional RANSAC cannot update the scale estimate in a video sequence with the value of the scale changing.

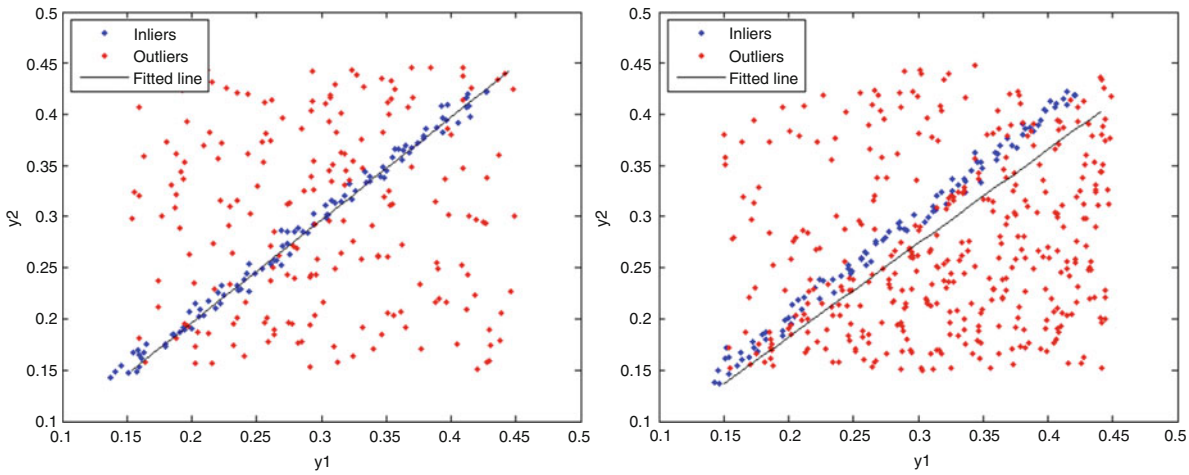
Over the past many years, several enhancements to the original RANSAC algorithm have been proposed. These enhancements fall into three main categories. The first kind deals with changes in the model verification. The second examines the way the sampling process generates hypotheses. The third category estimates from the data the scale of the inlier noise. The three types of enhancements were never attempted to improve RANSAC together.

In *model verification*, Matas and Chum used the $T_{d,d}$ test to find first a small set of inliers which they then extended to a larger set. In the *bailout* test, Capel assumed that a random subset of the data has a hypergeometric distribution for the inliers and tried to find a new hypothesis with a larger number of inliers after verifying it with the current subset. The procedure *WaldSAC*, described by Chum and Matas, used Wald's theory of sequential decisions to reject hypotheses that have the likelihood ratio of points belonging to a good and bad hypothesis less than a certain threshold.

These methods work correctly only if certain assumptions are satisfied. For example, the $T_{d,d}$ test may reject a correct hypothesis if the points in the initial subset are not all inliers, the bailout test could signal a false bailout if the lower bound on the number of inliers in the observed set of data points is not estimated correctly, and WaldSAC will be optimal only if the number of outliers are known a priori. Also, relative to the original RANSAC, for all these methods, often more parameters are required to be estimated from the data.

Two other methods were also developed that loosely belong to this category. In *preemptive* RANSAC, Nister started with a fixed number of hypotheses and used a breadth-first approach to evaluate and score all the hypotheses on a subset of the data in parallel. At each step, the hypotheses with best scores are evaluated on the new subset of data in the next iteration. As the method proceeds, the number of hypotheses decreases based on a preemption function, and the final estimate is returned by the procedure. In RANSAC for (quasi)-degenerate data (QDEGSAC) proposed by Frahm and Pollefeys, the method first detects possible degeneracies in the data and after that converges to the correct model.

Using *a priori* information introduces nonuniform sampling which can improve the quality of the generated hypotheses. In Progressive Sample Consensus (PROSAC), Chum and Matas used similarity scores of matching points to build a recurrence relation of the data points and used it to sample data points



Robust Methods, Fig. 2 Line fitting example using RANSAC. *Left*: the method works in case of fewer, symmetrical outliers. *Right*: it fails in the presence of a large number of unsymmetric outliers

nonuniformly thus converging to inliers faster. In *guided sampling*, Tordoff and Murray extended Torr and Zisserman's method of uniform sampling of the data points, called Maximum Likelihood Estimation Sample Consensus (MLESAC) [8]. They first computed matching probability of the feature points across two frames and then used these probabilities to sample the points nonuniformly. In locally optimized RANSAC (Lo-RANSAC), Chum et al. proposed generating a fixed number of *additional* hypotheses from the inliers already found by the algorithm and check against the whole dataset for improvement. Since the detected points are all inliers, there is no need to use the elemental subset for model estimation.

The reviews of all the algorithms above are given in [10]. Many of these variants are fast, and some can process data in real time. These methods are valid for the respective applications presented but will fail in general. Also, all these methods return only one inlier structure. Note that procedures which assume certain distribution for inliers or outliers will not be able to succeed in problems where the assumption may not hold.

The third category extracts, with additional processing, the *scale of inlier noise* from the data. The $k \times k$ scale matrix $\mathbf{S}$ is diagonal and is computed element by element. Some methods order the residuals assuming a Gaussian distribution for the inliers and take the variance of the k-th sample as an estimate of the scale. A more advanced scalar estimate uses

$$s = n^{-1/5} \underset{j}{\text{med}} \mid z_j - \underset{i}{\text{med}}\, z_i \mid, \tag{7}$$

where z_i is the projection of data points on to one of the k dimensions of $\mathbf{x}_i$.

Note that Eq. (7) differs from median absolute deviation (MAD) only by $n^{-1/5}$ and is symmetric relative to the median. Wang and Suter [11] employed mean shift mode and mean shift valley techniques using over-smoothed kernel bandwidth to estimate the scale of the inlier noise assuming the noise to be Gaussian. In the projection-based M-Estimators (pbM), Subbarao and Meer [12] computed for each hypothesis $\boldsymbol{\Theta}$, a new scale using the estimate of Eq. (7). Using mean shift for the intercept, the pbM usually recovered the right estimate. These extentions of RANSAC can recover three/four inlier structures but only under a quasi-symmetric outlier structures. In data with strongly asymmetric outliers, these methods will not work if there are more than one inlier structures.

The *generalized principal component analysis* (GPCA) introduced by Vidal et al. [13] is the last robust method that should be mentioned. The method can estimate an unknown number of inlier subspaces by analyzing the derivatives of polynomials obtained from data. Each of these subspaces can have different number of dimensions. The method is elegant, detects a few inlier structures in lower dimensions, but fails in the presence of unstructured outliers.

Application

In computer vision, robust methods have been applied to fundamental matrix estimation, trifocal tensor estimation, camera pose estimation and structure from motion. See [14] for references. Among all robust methods, RANSAC is topic of active research and newer enhancements to the original algorithm are proposed in all major vision conferences and journals. These papers will be judged positively only if the methods promise generalizability to large classes of problems involving images or image sequences.

Open Problems

A robust algorithm should be able to meet three fundamental goals in estimation. Firstly and most importantly, the scale of the inlier noise should be estimated automatically yet correctly, even if the outliers are strongly asymmetric. Secondly, the main estimation should retain from RANSAC the elemental subset-based hypothesize-and-test framework and from M-estimator the squared Mahalanobis distance with the diagonal scale matrix. Finally, the objective function should be chosen carefully taking into account the most general nature of the estimation problems with multiple inlier structures and small inlier to outlier ratios. Additionally, the heteroscedasticity of the linearized variables should be taken into consideration, and processing over Grassmann manifold using group theoretic methods should be an option [15].

References

1. Comaniciu D, Meer P (2002) Mean shift: a robust approach toward feature space analysis. IEEE Trans Pattern Anal Mach Intell 24(5):603–619
2. Kanatani K, Morris D (2001) Gauges and gauge transformations for uncertainty description of geometric structure with indeterminacy. IEEE Trans Inf Theory 47:2017–2028
3. Subbarao R, Meer P (2009) Nonlinear mean shift over Riemannian manifolds. Int J Comput Vis 84:1–20
4. Huber PJ (1996) Robust statistical procedures, 2nd edn. SIAM, Philadelphia
5. Rousseeuw PJ, Leroy AM (1987) Robust regression and outlier detection. Wiley, New York
6. Meer P, Mintz D, Kim DY, Rosenfeld A (1991) Robust regression methods in computer vision: a review. Int J Comput Vis 6:59–70
7. Stewart CV (1999) Robust parameter estimation in computer vision. SIAM Rev 41:513–537
8. Meer P, Stewart CV, Tyler DE (eds) (2000) Robust statistical techniques in image understanding. Special issue, Computer Vision and Image Understanding vol 78. Academic Press San Diego
9. Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 24(6):381–395
10. Raguram R, Frahm JM, Pollefeys M (2008) A comparative analysis of ransac techniques leading to adaptive real-time random sample consensus. In: ECCV '08: proceedings of the 10th European conference on computer vision, Springer, pp 500–513
11. Wang H, Suter D (2004) Robust adaptive-scale parametric model estimation for computer vision IEEE Trans Pattern Anal Mach Intell 26:1459–1474
12. Subbarao R, Meer P (2006) Beyond RANSAC: user independent robust regression. Workshop on 25 years of RANSAC, New York, NY
13. Vidal R, Ma Y, Sastry S (2005) Generalized principal component analysis (GPCA). IEEE Trans Pattern Anal Mach Intell 27:1945–1959
14. Hartley RI, Zisserman A (2004) Multiple view geometry in computer vision, 2nd edn. Cambridge University Press, Cambridge
15. Mittal S, Anand S, Meer P (2012) Generalized Projection Based M-Estimator. IEEE Trans Pattern Anal Mach Intell., 34:2351–2364

Robust Regression

▸Robust Methods