Chapter 2
Motion Analysis: Past, Present and Future

J.K. Aggarwal

Abstract The subject of motion has been the center of interdisciplinary studies since the time when Zeno posed his paradox circa 500BC. However, computer vision, the use of a camera and a computer to recognize objects, people and/or events automatically, is a relatively young field of research. Its development began in the early 1960s; however, it has matured fairly quickly. Today, it is contributing to the solutions of some of the most serious societal problems. Motion analysis of a sequence of images is an important part of computer vision. This chapter briefly presents the contributions to motion analysis from other fields followed by the computer vision-based analysis of motion from a sequence of images. Analysis and understanding of images based on both feature tracking and optical flow estimation are presented. Early works focused on the computation of structure from motion of objects from a sequence of images via point features. This was followed by the computation of optical flow to characterize motion. Applications today focus on the monitoring of traffic, providing guidance to a motorist in terms of his/her position relative to traffic lanes and traffic ahead, and inspection of complicated three-dimensional industrial parts, to mention a few. Research focus has shifted from inanimate objects to people, for example monitoring people and their activities in public places or monitoring activities from an unmanned aerial vehicle. These applications are dominating the research scene through the belief that computer vision/motion analysis can contribute to the solution of societal surveillance and biometric problems. The chapter ends with a discussion of the future directions of research in motion analysis and possible applications.

Keywords Motion · Tracking · Human activities · Past · Present · Future

J.K. Aggarwal (✉)
The University of Texas at Austin, Austin, TX, USA
e-mail: aggarwaljk@mail.utexas.edu

B. Bhanu et al. (eds.), *Distributed Video Sensor Networks*,
DOI 10.1007/978-0-85729-127-1_2,

1 Introduction to Motion: An Early History

Timely response to moving objects is of critical importance; at times, it is a matter of life and death. It is natural that the mammalian visual system has evolved specialized neural hardware to detect and respond to an ever-changing environment full of stationary objects, moving objects, and objects that appear to move due to the movement of our bodies or eyes. Thus, it is not surprising that human interest in motion as a subject of scientific thought emerged long ago. This interest has spanned diverse disciplines including philosophy, psychophysics, neurobiology, robotics, computer graphics, and, of course, computer vision.

Motion and human vision has been the subject of intense inquiry both at experimental and theoretical levels in computer vision research. Interest in computer vision itself began in the early 1960s and research in computer vision-based analysis and understanding of motion in a sequence of images began around the early 1970s. The publication of the paper by Johansson [23] provided significant impetus for the study of motion in several diverse disciplines. Author's interest in motion analysis began around that time. The paper "Computer Analysis of Moving Polygonal Images," in *IEEE Transactions on Computers* [4] is one of the earliest papers published on motion in computer vision literature. This paper was motivated by the desire to track objects; more details follow in Sect. 3. Author's interest in tracking three-dimensional objects, and Norman Badler's interest in movements of dance led to the 1979 Workshop on Computer Analysis of Time Varying Imagery in Philadelphia. It was one of the first meetings devoted to research on machine perception of motion [7]. A subsequent workshop of note includes the 1982 NATO Advanced Study Institute on Image Sequence Processing and Dynamic Scene Analysis directed by T.S. Huang [22]. There were several other workshops sponsored by the IEEE Computer Society and by the Association for Computing Machines (ACM). In addition, several other workshops were held outside the USA, for example those chaired by Vito Cappellini in Florence, Italy.

Shimon Ullman at MIT submitted his dissertation in 1977 on the motion of objects [38]. The 1980s were a very exciting period for research in motion analysis and understanding. Berthold Horn (with Brian Schunck) published his seminal paper on Optical Flow in 1981 [21] and published his book on *Robot Vision* in 1986 [20]. In 1988, Worthy Martin and the author edited a collection of papers on various facets of motion entitled *Motion Understanding: Robot and Human Vision* [26]. The authors included Shimon Ullman, Thomas Huang, Hans Nagel, and David Fleet and others. In addition, several applications, which are the lifeblood of any research, were also presented in the book. Today, the diverse applications motivating motion research include medicine, tomography, autonomous navigation, communications, television, video-conferencing, unmanned aerial vehicle imagery, athletics, dance choreography, and meteorology.

Given the constraints of space and time, it is difficult to cover all aspects of motion analysis and understanding. In this chapter, the author presents an overview of the past, present and future of research in Motion and related issues. In Sect. 2, highlights of motion research from other disciplines are briefly mentioned. Section 3 covers some of the beginnings of computer vision-based research in motion.

Section 4 is devoted to Optical Flow. Section 5 describes studies in recognition of human activities. Section 6 presents a view of the future directions of research and applications.

2 Motion: Highlights from Philosophy, Psychology and Neurobiology

Understanding motion, the movement of objects through time and space, has interested humans for at least 2500 years. In the 5$^{\text{th}}$ century BC, the Greek philosopher Zeno of Elea raised the first major scientific question to inquire regarding the nature of motion: Is it discrete or continuous? He presented the "paradox of arrow" to point out the logical problem encountered when motion is conceived to be the repositioning of an object at successive time instances in a given spatial coordinate system. While some modern philosophers like Bertrand Russell [33], and certain aspects of theory of special relativity as detailed on the Math Pages website [27] have "resolved" Zeno's paradox, the point remains that the phenomenon of motion has played a long and pivotal role in the evolution of scientific thought.

Aristotle, more than two millennia ago, posed questions about the nature of motion. He made one of the first references to a motion illusion: the motion after effect. The motion after effect is created when the human visual system receives a prolonged stimulation from a moving field in a particular direction like a swift river or a waterfall. When one views a stationary object after one's vision has accommodated to the motion field, the stationary object appears to move in the opposite direction. This illusion was rediscovered by Addams in 1834 [1]. While Aristotle talks about a swiftly moving river, Addams mentions a waterfall.

The human vision system and its relationship to motion have been studied intensely by psychophysicists of various persuasions. Based on the findings presented in the excellent review article by Derrington et al. [12], one may observe that there are two distinct visual subsystems that analyze retinal motion, a feature-tracking subsystem and a motion-sensing subsystem.[1] Paraphrasing Derrington et al.:

> A *feature-tracking* motion subsystem tracks features or objects across multiple frames.
>
> A *motion-sensing* subsystem consists of two components:
>
> A "*first order*" motion filtering system is the substrate of our direct sense of motion. It consists of local orientation-selective motion sensors that respond to the motion of luminance-defined features in a particular direction. The location, speed and direction of motion of a moving object are encoded in the visual system by the identity of the sensor that responds to it.
>
> A "*second order*" motion system is based on sensors that respond to motion of patterns defined by spatial modulation of local contrast or other image properties.

[1]The interested reader may note that the first physiological model of motion detection based on the "Autocorrelation, a principle for evaluation of sensory information by nervous system" by Reichardt W. (1961) is presented at the website: http://en.wikipedia.org/wiki/Motion_perception.

The nature of models used for motion detectors in the visual system has been the subject of considerable thought and research. The most commonly accepted models are by Adelson and Bergen [2] and Watson and Ahumada [42]. There is no clear agreement on how the above mechanisms give rise to vection and optical flow. Also, there is disagreement on the mode of operation of the motion filters and how their outputs are combined to resolve the direction of 2D motion. Moreover, it is difficult to specify the differences in performance between the feature-tracking mechanisms and the motion-sensing mechanisms.

Another class of study that addresses the visual perception of motion is in the field of neurobiology. The work of Semir Zeki [45] led to the acceptance of the concept of functional specialization in the visual cortex, where color, form, and motion are processed separately. Four parallel systems handle different attributes of vision: one for motion, one for color, and two for form. It is important to note that the two channels that are computationally most distinct from one another are motion and color systems. For the human visual-motion system, the pivotal area of the brain is called V5, a lesion in this area produces akinetopsia where patients neither see nor understand the world in motion. At rest, objects appear perfectly visible, whereas motion relative to them causes the object to vanish.

Motion and human understanding of motion have been studied by a number of different disciplines. Motion is a fascinating subject with a number of interesting unsolved problems. This fascination with motion and the human visual system is likely to continue far into the future!

3 Motion in Computer Vision: The Beginnings

Having presented some of the highlights from philosophy, psychology and neurobiology, let us consider the beginnings of motion studies in computer vision. The problem of determining cloud motion from a sequence of satellite images motivated the paper [4]. The difficulty of approximating clouds and cloud motions led to the idealization of polygonal figures moving in planes. The assumed model of cloud motion consists of polygons at different levels moving independently and the observed view of cloud motion is the overlapping view of various polygonal figures at various layers. The polygons are assumed to be rigid in order to determine the linear and angular velocities of the polygons, and to decompose the sequence of images into component sequences of images. In reality, clouds are neither polygonal nor rigid; indeed, clouds change shape, appear and disappear. Nevertheless, the study revealed a number of interesting and fundamental relations concerning overlaying polygonal figures. The study yielded one program to generate a sequence of images from the moving polygons in different layers and another program (analysis program) to systematically analyze the generated overlay of polygonal image sequences. The analysis program correctly constructed the description of the underlying polygons. The features used in tracking polygons included vertices, angles subtended at the vertices and the length of sides of polygons.

The above scenario is similar to a person observing a scene through a narrow slit. Let us assume that a car passes by the slit. At no time does the observer see the car in its entirety but the observer is able to construct a description of the car and determine its speed. Our motivation for studying two-dimensional motion of objects in a plane was a matter of convenience and it was a meaningful first step. Further, the use of simple polygonal shapes was also driven by a lack of digitized images, digitizing hardware for video and sufficiently powerful computing equipment. However, the resulting ability to 'understand' the sequence of polygonal images led to more realistic scenarios.

In general, researchers are interested in three-dimensional (3D) objects in three-dimensional space. However, psychologists have reported several cases of humans discerning a 3D structure from moving 2D images. Wallach and O'Connell [41] presented silhouettes of moving wire-frame objects to subjects and the subjects reported seeing a three-dimensional structure while individually the frames contained no information about the three-dimensional structure. Further, under the right circumstances the observer could describe the wire-frame structure correctly. Wallach and O'Connell called it the "kinetic depth effect". Shimon Ullman [39] called this the "structure from motion" effect. He considered the estimation of a three-dimensional structure from moving light displays, i.e. images formed by the orthographic projection of rigid configurations of localized feature points. Ullman derived the result that four non-coplanar points in three distinct views are sufficient to give the three-dimensional structure of the points. John Roach [32] considered the same problem but with central projection (or perspective projection). The problem here was significantly more complex because the determination of the structure depended upon the solution to a set of nonlinear equations. Roach[2] discovered that two views of six points led to an over-determined set of equations. Roach also observed that one may use five points but the solution was not always stable.

This is an example of feature tracking which corresponds to the first subsystem (of the human visual system) mentioned in Sect. 2. In this case, the points are presumably extracted from features on moving objects of interest. The underlying assumption is that the corresponding points represent the feature points on the moving object. In the above analysis, features played an important role in computing motion and regions of 'similar' motion.

Another important contribution was made by Jon Webb [43], inspired by Johansson's [23] experiments in which Johansson showed his subjects a movie of a person walking around in a dark room with lights attached to the person's major joints. Even though only lights could be seen, there was a strong impression of three-dimensional motion in these movies. In these experiments, each rigid-body part was represented by two points. Webb considered the motion of two rigidly connected points and observed that any motion may be decomposed into a rotation and

[2]The methods for solving nonlinear equations have progressed significantly, as well as the available computing power, so that now there is a software package, the EOS Systems Photomodeler [13], which uses two views of six points to generate a structure. The software package first appeared on the market in 1999, indicating the long interval between an idea and a product.

a translation, however, Webb made an additional assumption that the axis of rotation is fixed in a direction for a short period of time. Webb was able to exploit this to determine the structure of jointed objects under orthographic projections. Webb applied his algorithm to real data with excellent results. This is one of the earliest studies on the motion of jointed objects and the analysis of human motion.

In the works described above, feature points are used to estimate motion as well as the structure of objects. There are two difficulties with this paradigm. First, the correspondence between points in successive images must be established, and second, one must take into account the disappearance of points due to occlusion and possibly noise. The establishing of correspondence is a serious problem even in simple environments. An alternative way of computing motion follows where correspondence is established between regions of the image plane as the first step.

4 Optical Flow-Based Motion Detection

It is mentioned in Sect. 2 that the human visual system responds to motion stimulus in mainly two ways: feature-based detection of motion and through a process driven by direction-of-motion sensors, even though it is not known how the motion sensors give rise to optical flow. Section 3 above described feature-based sensing and discussed briefly the advantages and disadvantages of feature tracking and sensing motion based on features. In this section, the beginnings of the optical-flow based techniques are presented.

The motion of an object in three-dimensional space when imaged by a camera system yields a vector field of the motion vectors corresponding to each point of the object. It is assumed that the image surface is a plane and the vector field is constructed by observing the position of a given point on the object at two successive instants of time. The duration of the time interval between the two instants of time may be made arbitrarily small. At any given time, the direction of the flow vector at a specific point on the object is the projection of the direction of motion of the point in space, and the magnitude of the vector is the projection of the length of the velocity vector of the object point. This flow field is called the optical-flow field. Thus, optical flow is a mapping of the "apparent" velocities of points on the object as projected on the image plane. This distribution of velocities can provide a wealth of information including motion of objects, presence of multiple objects, and rate of movement of objects. It has been the subject of intense study by computer-vision researchers.

A fundamental contribution is due to Horn and Schunck [21]. The brightness of an object point, $f(x, y)$, may be assumed to be constant in a small window for the purposes of computing optical flow. They formulated the problem using (1) as follows:

$$\frac{df(x,y)}{dt} = 0. \tag{1}$$

Equation (1) may be expanded as:

$$\frac{\partial f(x,y)}{\partial x}\frac{dx}{dt}+\frac{\partial f(x,y)}{\partial x}\frac{dy}{dt}+\frac{\partial f}{\partial t}=0 \tag{2}$$

or

$$\frac{\partial f(x,y)}{\partial x}u+\frac{\partial f(x,y)}{\partial x}v+\frac{\partial f}{\partial t}=0, \tag{2a}$$

where $u=\frac{dx}{dt}$ and $v=\frac{dy}{dt}$.

Here f is the image intensity at point (x, y) in the image plane and (u, v) is the vector describing the velocity vector and ∂ is the partial derivative operator. This is only one equation. Obviously, (1) is not adequate by itself to solve for the velocity vector (u, v). Additional constraints must be imposed. A variety of constraints may be imposed to solve this equation. Horn and Shunck suggested two constraints as follows:

(i) By minimizing the square of the magnitude of the gradient of the optical-flow velocity

$$\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u}{\partial y}\right)^2 \quad \text{and} \quad \left(\frac{\partial v}{\partial x}\right)^2+\left(\frac{\partial v}{\partial y}\right)^2. \tag{3}$$

(ii) Or by minimizing the sum of the squares of the Laplacian of x- and y-components of flow

$$\nabla^2 u=\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2} \quad \text{and} \quad \nabla^2 v=\frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2}. \tag{4}$$

Today there are a number of techniques for computing optical flow. Fleet and Weiss [14] present a review of methods for the computation of optical flow. The computation of optical flow is important in many applications and the method by Ogale and Aloimonos [29] is particularly useful.

Lucas and Kanade [25] presented an iterative technique for image registration with application to stereo vision.[3] Later, Tomasi and Kanade [36] adapted it for detection and matching of feature points. The proposed two frame method is probably the most commonly used method for computing optical flow. Optical flow is based on image patches and an affine model for the flow field. Essentially the flow is assumed to be locally constant.

At a more theoretical level, several researchers are pursuing more elaborate methods for computing optical-flow and motion boundaries. In particular, the methodology introduced by Mumford and Shah [28] for computation of homogeneous regions and smooth boundaries by minimizing a functional has become a tool of choice. Deriche et al. [11] document results on optical-flow estimation based on

[3]Professor Takeo Kanade once remarked to the author at a conference that the complexity of the Roach and Aggarwal [32] solution to the feature-tracking problem contributed to the Lucas and Kanade [25] simple solution. Further, the Lucas and Kanade solution is probably his most contributive paper among all his contributions to date.

a variational approach and Vazquez et al. [40] present results on region and motion estimation. This is an active area of research at present. These methods will probably grow further and be incorporated in the solution of real world problems in the near future.

5 Human Actions and Activities

In the above two sections a discussion of detection of motion paralleling the physiological model for detection of motion has been presented. There are several other methodologies for detection of motion in computer vision. Instead of focusing attention on just detection of motion, an application of computer vision, the recognition of human actions and activities, will be considered in this section. Recently, computer vision-based recognition of human actions and continued activities has gained significant importance in view of the intense desire for safety, security and surveillance in almost every aspect of life. The surveillance of people at public places like airports and subway stations is well known, however, there are a number of other applications that make the observation of persons in normal activities at home and in the workplace as interesting and as challenging. Several funding agencies have initiated substantial programs to study human activities in various forms including observation of human activities from unmanned aerial vehicle based cameras. Monitoring the elderly in a 'smart' home equipped with multiple cameras and other sensors and alarms, analysis and understanding of sports video, and the content-based video summarization and retrieval (useful for video sharing websites) are several different types of applications. Another application involves the monitoring of industrial environments where humans perform hazardous tasks wearing important protective clothing. In this application the safety of the product and the safety of persons are at stake. The movie industry is interested in synthesizing a given person's actions, gait and other characteristics based on a model video. In military applications there is an added constraint of real-time analysis and understanding, thus, there is a plethora of applications where computer vision-based recognition of human activities may contribute to society.

Before one is able to understand an activity based on a sequence of images, many low-level image processing steps in along chain of steps must be performed. The motion analysis methods presented in Sects. 3 and 4 serve as low-level components for activity-level analysis. The efficacy of these steps determines the overall success of determining high-level activities. A look at real imagery for a given application, for example imagery in subway station, requires careful algorithms for segmentation and extraction of features. In addition, surveillance and certain other applications, monitoring is a 24/7 problem including nighttime, rain and fog. Low-level issues are not discussed here but this is not an attempt to slight them. They must be dealt with and the results at higher levels depend on the efficacy of the tools at the low level.

Yamato et al. [44] introduced recognizing human actions in time-sequential images using hidden Markov models (HMMs). However, significant human activity recognition had its origin at the MIT Media Lab in Alex Pentland's group. Starner and Pentland [35] document early work on American Sign Language detection using HMMs. The paper by Oliver et al. [30] is another important paper delineating the segmentation of human blobs, estimating their position and velocity, and recognizing two-person interactions by applying coupled HMMs. Several other works appeared at about the same time including Haritaoglu et al. [19], Haritaoglu and Davis [18] and Cai and Aggarwal [8]. (The group at The University of Texas at Austin derived much inspiration from these works.) It is worth emphasizing that a person is just a blob in the above works. Later, Park and Aggarwal [31] introduced the motion of body parts to recognize human actions. This was a distinct departure from the earlier paradigm. Other works with similar flavor include Gavrila and Davis [17] and Campbell and Bobick [9].

Most of the above works focus on simple and short duration activities. Michael Ryoo [34] presents a context-free-grammar based representation scheme as a formal syntax for representing composite and continued activities. The system is able to recognize recursive activities such as assault and fighting. The approach is able to recognize continued and crowd/group activities.

The period of late 1990s and early 2000s has had an enormous amount of research activity in the area of human actions and activities. The earlier reviews by Cedras and Shah [10], Aggarwal and Cai [3] and Gavrila [16] document the research prior to 2000. The more recent reviews by Turga et al. [37] and Aggarwal and Ryoo [6] detail the recent human activity research.

An approach based taxonomy is outlined in the review by Aggarwal and Ryoo [6]. In particular, the methodologies are classified as single layered approaches and multi-layered approaches. Single layered approaches are more suitable for recognizing simple actions whereas multi-layered approaches are more relevant for complex or continued human activities. Single layered approaches may be further subdivided into space-time approaches or sequential approaches. Multi-layered or hierarchical approaches may also be further classified as statistical, syntactical or description-based approaches. Statistical approaches use Hidden Markov Models, syntactic approaches use grammar syntax such as stochastic context-free grammars, and description-based approaches describe events in terms of sub-events of actions and their temporal, spatial and logical structure. The last approach is particularly flexible and is to capture the variability of human activities.

The author believes that in the not too distant future significant results on the recognition of human actions and activities will lead to products in industrial and military domains.

6 Motion: Future

Obviously, we have made significant progress in developing methodologies and algorithms for analyzing and understanding sequences of images. In addition, we have made significant progress because computation is faster, memory is cheap, and cameras are relatively inexpensive. In the 1970s, it used to take a VAX 11/780 to process images and the cost of a respectable system was about $100,000, whereas today one can build a very good computer vision system with $2,000. Further, the ability to acquire and process images fairly effortlessly has contributed to the development of a large number of applications.

However, the author would be remiss without mentioning that problems are difficult and solutions are incomplete. A presentation at a computer vision workshop entitled "Structure and Motion From Images: Fact and Fiction" [5] outlined the difficulties by using points, lines, and contours as discrete features and optical-flow and range images as additional features to analyze motion. It also presented difficulties arising from occlusion, low-level processing, and time reliability constraints. Paraphrasing the conclusion in 1985, the same holds today "... we have performed fairly but much remains to be resolved."

Computer vision-based research in motion is already contributing to some of the pressing societal problems. In particular several corporations have products. Siemens has installed a traffic monitoring system to monitor traffic in tunnels. It is able to estimate the level of traffic and signal to motorists. It has been installed and tested in a variety of conditions. Another system, manufactured by MobilEye, installed inside an automobile, warns the driver of an impending collision or if the car is going outside the prescribed lane. It is also available as a product. National Instruments is producing an inspection system for automobile parts and this particular system inspects a complicated part by measuring its dimensions in a three- dimensional coordinate system. The part has to satisfy very precise dimensions. There are systems being deployed at airports to detect the direction of the movement of passengers. So it is fair to say that computer vision is producing systems for the well being of society.

The future of computer vision-based research in motion is bright. One factor contributing to rapid progress is the tools that are coming on line. The "computing scene" including speed and packaging is improving by the day! The price of memory is going down and the resolution (definition) of cameras has increased substantially. The research of Shree Nayar [24] at Columbia University and of Narendra Ahuja [15] at The University of Illinois, Urbana/Champaign is contributing to the developments of new camera technologies. Embedded cameras are available with a significant amount of computing power and more innovations in camera technology are on the way. These cameras are becoming relatively cheap. In addition to intensity image cameras, cameras with other modalities are finding use in daily life. Image processing and understanding using other modalities offers new avenues of investigation. Robot vision with understanding of motion and environment poses the greatest challenges and rewards. The success of the MARS Rover has brought forth a lot of interest in computer-vision driven robot exploration. With real-time constraints, the problems are truly challenging.

It is a given that every cell phone has a camera and there are billions of cell phones in the world today. In addition, people are carrying them everywhere they go. It may not be too far-fetched to think that cameras will be used for navigation and monitoring. That is clearly a great application of cell phone and computer vision.

In the not too distant future, the following scenario of a synthetic movie of a person appears likely. We observe a person in various poses and activities and we are able to construct a repertoire of motions of the person. Based on the analysis of these motions, we synthesize images of the person in any given posture and possibly a motion picture of that person performing various activities not included in the repertoire. In this way, we will be able to synthesize movies of persons in their absence. This will be a true interplay between computer vision and computer graphics.

A number of researchers are pursuing the recognition of everyday human activities when observed from an unmanned aerial vehicle. The difficulties of imaging noise, stabilizing images and a host of other problems make the problem very difficult at this stage. Problems of monitoring workers in the food industry or workers performing other hazardous tasks, where protective gloves and other clothing are necessary, are beginning to receive attention. Monitoring patients in a hospital environment is another area of application. Given all the problems begging for a solution, the future is truly bright for motion research and applications.

Motion is a subject of serious study in computer vision, image processing, computer graphics, and robot vision communities. The objectives are different in each of the communities in view of their applications. For example: in computer vision one is interested in human activities; in image processing video compression and transmission take the front seat; in computer graphics the production of life-like synthetic images takes precedence; and in robot vision real-time navigation applications dominate. The future of research and applications in motion in each of these endeavors is paved with low hanging fruit.

Acknowledgements It is a pleasure to acknowledge the help and comments of Dr. Michael Ryoo, Professors Bill Geisler and Amar Mitiche and Mr. Birgi Tamersoy and Mr. Chia-Chih Chen. Also, my sincere thanks go to Ms. Selina Keilani for editing the manuscript. The research was supported in part by Texas Higher Education Coordinating Board award # 003658-0140-2007.

References

1. Addams, R.: An account of a peculiar optical phenomenon seen after having looked at a moving body, etc. Lond. Edinb. Philos. Mag. J. Sci., 3rd Ser **5**, 373–374 (1834)
2. Adelson, E.H., Bergen, J.R.: Spatiotemporal energy model for the perception of motion. J. Opt. Soc. Am. A **2**, 284–299 (1985)
3. Aggarwal, J.K., Cai, Q.: Human motion analysis: a review. Comput. Vis. Image Underst. **73**(3), 428–440 (1999)
4. Aggarwal, J.K., Duda, R.O.: Computer analysis of moving polygon images. IEEE Trans. Comput. **24**(10), 966–976 (1975)
5. Aggarwal, J.K., Mitiche, A.: Structure and motion from images: fact and fiction. In: Proceedings of the Third Workshop on Computer Vision: Representation and Control, Bellaire, MI (1985)
6. Aggarwal, J.K., Ryoo, M.S.: Human activity analysis: a review. ACM Comput. Surv. (2010, to appear)
7. Badler, N.I., Aggarwal, J.K.: Abstract of the Workshop on Computer Analysis of Time-Varying Imagery, Philadelphia, PA (1979)
8. Cai, Q., Aggarwal, J.K.: Tracking human motion in structured environments using a distributed camera system. IEEE Trans. Pattern Anal. Mach. Intell. **21**(12), 1241–1247 (1999)
9. Campbell, L.W., Bobick, A.F.: Recognition of human body motion using phase space constraints. In: IEEE Int. Conference on Computer Vision (ICCV), pp. 624–630 (1995)

10. Cedras, C., Shah, M.: A motion-based recognition: a survey. Image Vis. Comput. **13**(2), 129–155 (1995)
11. Deriche, R., Kornprobst, P., Aubert, G.: Optical-flow estimation while preserving its discontinuities: a variational approach. In: Proceedings Second Asian Conference on Computer Vision. Lecture Notes in Computer Science, vol. 1035, pp. 71–80 (1995)
12. Derrington, A.M., Allen, H.A., Delicato, L.S.: Visual mechanisms of motion analysis and motion perception. Annu. Rev. Psychol. **55**, 181–205 (2004)
13. Eos Systems, Inc. Photomodeler, http://www.photomodeler.com/
14. Fleet, D.J., Weiss, Y.: Optical flow estimation. In: Paragios, et al. (eds.) Handbook of Mathematical Models in Computer Vision. Springer, Berlin (2006)
15. Gao, C., Hua, H., Ahuja, N.: A hemispherical imaging camera. Comput. Vis. Image Underst. **114**(2), 168–178 (2010)
16. Gavrila, D.: The visual analysis of human movement: a survey. Comput. Vis. Image Underst. **73**(1), 82–98 (1999)
17. Gavrila, D., Davis, L.: Towards 3-D model-based tracking and recognition of human movement. In: Bichsel, M. (ed.) International Workshop on Face and Gesture Recognition, pp. 272–277 (1995)
18. Haritaoglu, I., Davis, L.: Hydra: multiple detection and tracking silhouettes. In: IEEE Workshop on Visual Surveillance, Fort Collins, CO, pp. 6–13 (1999)
19. Haritaoglu, I., Harwood, D., Davis, L.: W4: who, when, where, what: a real system for detecting and tracking people. In: Third Int. Conf. on Automatic Face and Gesture Recognition, Nara, Japan, pp. 222–227 (1998)
20. Horn, B.K.P.: Robot Vision. MIT Press, Cambridge (1986)
21. Horn, B.K.P., Schunck, B.G.: Determining optical flow. Artif. Intell. **17**, 185–203 (1981)
22. Huang, T.S.: Image Sequence Processing and Dynamic Scene Analysis. NATO ASI Series. Springer, Berlin (1983)
23. Johansson, G.: Visual motion perception. Sci. Am. **232**(6), 76–88 (1975)
24. Krishnan, G., Nayar, S.K.: Cata-fisheye camera for panoramic imaging. In: IEEE Workshop on Application of Computer Vision (WACV) (2008)
25. Lucas, B.D., Kanade, T.: An iterative image registration technique with application to stereo vision. In: Proceedings of 7th International Joint Conference on Artificial Intelligence (1981)
26. Martin, W.N., Aggarwal, J.K. (eds.): Motion Understanding: Robot and Human Vision. Kluwer Academic, Dordrecht (1988)
27. Math Pages Website (http://www.mathpages.com/rr/s3-07/3-07.htm), Reflections on relativity—Zeno and the paradox of motion
28. Mumford, D.S.: Boundary detection by minimizing functionals. In: Proceedings of CVPR (1985)
29. Ogale, A.S., Aloimonos, Y.: A roadmap to the integration of early visual modules. Int. J. Comput. Vis. **72**(1), 9–25 (2007). Special Issue on Early Cognitive Vision
30. Oliver, N., Rosario, B., Pentland, A.: A Bayesian computer vision system for modeling human interactions. In: Proceedings of International Conference on Vision Systems 99, Gran Canaria, Spain, pp. 255–272 (1999)
31. Park, S., Aggarwal, J.K.: A hierarchical Bayesian network for event recognition of human actions and interactions. Multimed. Syst. **10**(2), 164–179 (2004)
32. Roach, J.W., Aggarwal, J.K.: Determining the movement of objects from a sequence of images. IEEE Trans. Pattern Anal. Mach. Intell. **2**(6), 554–562 (1980)
33. Russell, B.: The problem of infinity considered historically. In: Salmon, W.C. (ed.) Zeno's Paradoxes, p. 51. Bobbs-Merrill, Indianapolis (1970)
34. Ryoo, M.S., Aggarwal, J.K.: Semantic representation and recognition of continued and recursive human activities. Int. J. Comput. Vis. **82**, 1–24 (2009)
35. Starner, T., Pentland, A.: Real-time American sign language recognition from video using hidden Markov models. In: International Symposium on Computer Vision, p. 265 (1995)
36. Tomasi, C., Kanade, T.: Detection and tracking of point feature shape and motion from image streams: a factorization method—part 3. Technical Report CMU-CS91-132 (1991)

37. Turga, P., Chellappa, R., Subrahmanian, V.S., Udren, O.: Machine recognition of human activities: a survey. IEEE Trans. Circuits Syst. Video Technol. **18**(11), 1473–1488 (2008)
38. Ullman, S.: The Interpretation of Visual Motion. MIT Press, Cambridge (1979)
39. Ullman, S.: The interpretation of structure from motion. Proc. R. Soc. Lond. B **203**, 405–426 (1979)
40. Vazquez, C., Mitiche, A., Langaniere, R.: Joint segmentation and parametric estimation of image motion by basis function representation and level set evolution. IEEE Trans. Pattern Anal. Mach. Intell. **28**(5), 782–793 (2006)
41. Wallach, H., O'Connell, D.C.: The kinetic depth effect. J. Exp. Psychol. **45**(4), 205–217 (1953)
42. Watson, A.B., Ahumada, A.J.: Model of human visual-motion sensing. J. Opt. Soc. Am. A **2**(2), 322–342 (1985)
43. Webb, J.A., Aggarwal, J.K.: Structure from motion of rigid and jointed objects. Artif. Intell. **19**, 107–130 (1982)
44. Yamato, J., Ohya, J., Ishii, K.: Recognizing human action in time-sequential images using hidden Markov model. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 379–385 (1992)
45. Zeki, S.: The visual image in mind and brain. Sci. Am. 69–76 (1992)

http://www.springer.com/978-0-85729-126-4

Distributed Video Sensor Networks
Bhanu, B.; Ravishankar, C.V.; Roy-Chowdhury, A.K.;
Aghajan, H.; Terzopoulos, D. (Eds.)
2011, XVII, 485 p., Hardcover
ISBN: 978-0-85729-126-4