Asynchronous Event Based Vision: Algorithms and Applications to Microrobotics

Zhenjiang Ni

INSTITUT DES SYSTEMES INTELLIGENTS ET DE ROBOTIQUE

UNIVERSITE PIERRE ET MARIE CURIE

P H D T H E S I S

Zhenjiang NI

Asynchronous Event Based Vision: Algorithms and Applications to Microrobotics

Thesis Advisor: Stéphane RÉGNIER

defended on Nov 26, 2013

Jury :

Yu SUN	-	Professor at University of Toronto	Reviewer
Sylvain SAÏGHI	-	Associate Professor at Université Bordeaux 1	Reviewer
Nathalie WESTBROOK	-	Professor at Institut d'Optique	Examinatrice
Michaël GAUTHIER	-	CNRS Senior Scientist at Université de Franche-Comté	Examinator
Stéphane RÉGNIER	-	Professor at Université Pierre et Marie Curie	Examinator
Ryad BENOSMAN	-	Associate Professor at Université Pierre et Marie Curie	Examinator

Asynchronous Event Based Vision: Algorithms and Applications to Microrobotics

Abstract: Although high speed camera technologies boost the image acquisition speed by orders, the image processing speed stagnates. The emerging discipline of neuromorphic engineering aiming to integrate the behaviours of living neural systems into large scale computer hardware or software breaks the bottleneck. The dynamic vision sensor (DVS) is a neuromorphic silicon retina prototype that records only scene contrast changes in the form of stream of events, thus naturally excluding the redundant background absolute gray levels. In this context, numerous high speed asynchronous event based vision algorithms have been developed and their advantages over frame based processing methods have been compared.

Real-world robotic applications are real-time and usually require fast sensory feedback. In haptic feedback teleoperated micromanipulation, vision is a sound candidate for force estimation if the position-force model is well established. The sampling frequency, however, needs to attain 1 kHz to allow a transparent and reliable tactile sensation and to ensure system stability, which is beyond the speed of the conventional frame based processing. The event based vision has thus been applied to provide the needed force feedback for two micromanipulation applications:

- Haptic feedback teleoperation of optical tweezers;
- Haptic virtual assistance in microgripper based micromanipulation.

The results show that the haptic frequency requirement of 1 kHz has successfully been achieved. For the first application, high speed particle position detection algorithms have been developed and validated. A three-dimensional haptic feedback system capable of manipulating multiple-trap optical tweezers has been realized. Extensive experiments have been conducted, of which a high quality shape sensation of a complex three-dimensional micro-object has been performed thanks to the high speed haptic feedback. In the second application, a novel event based shape registration algorithm capable of tracking arbitrary form object has been developed to track a piezoelectric microgripper in picking and placing a glass sphere with diameter 50 µm. The stability of the system has been significantly enhanced to intuitively assist operators in performing complex micromanipulation tasks.

Keywords: Dynamic Vision Sensor; Asynchronous Event Based Computation; Haptic Feedback; Teleoperation; Optical Tweezers; Microrobotics; Micromanipulation;

Vision Asynchrone Événementielle: Algorithmes et Applications à la Microrobotique

Résumé: Bien que les technologies de caméras rapide augmentent les vitesses d'acquisition par ordres, la vitesse de traitement traîne. La discipline émergente de l'ingénierie neuromorphique visant à intégrer les comportements des systèmes neuronaux en matériel ou logiciel informatique à grande échelle surmonte ce facteur limitant. Le Dynamic Vision Sensor (DVS) est un prototype de la rétine silicium neuromorphique qui n'enregistre que des changements de contraste de la scène sous la forme de flux d'événements, excluant donc naturellement les niveaux de gris absolus redondants. Dans ce contexte, de nombreux algorithmes de vision asynchrones à grande vitesse basées sur l'événement ont été développés et leurs avantages par rapport aux méthodes de traitement traditionnel basé sur l'image ont été comparés.

Les applications robotiques sont en temps réel et nécessitent généralement un retour sensoriel rapide. En retour haptique pour la micromanipulation, la vision est un candidat qualifié pour l'estimation de la force si le modèle position-force est bien établi. La fréquence d'échantillonnage doit toutefois atteindre 1 kHz pour permettre une sensation tactile transparente et fiable et assurer la stabilité du système, qui est au-delà de la vitesse du traitement classique basé sur l'mage. La vision basée sur l'événement a donc été appliquée pour fournir le retour d'effort nécessaire sur deux applications de micromanipulation :

- Le retour haptique sur la pince optique;
- Assistance virtuelle haptique sur micro-outil mécanique.

Les résultats montrent que l'exigence de fréquence haptique de 1 kHz a été réalisée avec succès. Pour la première application, les algorithmes de détection de la position d'une microsphère à haute vitesse ont été développés. Un système de retour haptique tridimensionnel capable de manipuler plusieurs pièges optiques a été réalisé. Des expériences approfondies ont été menées, dont une sensation de forme de haute qualité d'un micro-objet tridimensionnel complexe a été réalisée grâce au retour haptique à haute vitesse. Dans la deuxième application, un nouvel algorithme d'enregistrement de forme basé sur l'événement capable de suivre objet de forme arbitraire a été développé pour suivre une micropince piézoélectrique pendant le pick-and-place d'une sphère de verre d'un diamètre de 50 µm . La stabilité du système a été considérablement renforcée pour aider intuitivement les opérateurs à effectuer des tâches de micromanipulation complexes.

Mots clés: Capteur Vision Dynamique; Calcul Asynchrone Basé sur Événements; Retour Haptique; Téléopération; Pince Optique; Microrobotique; Micromanipulation;

Acknowledgements

As a PhD candidate, the most meaningful competence that I have acquired during this period is concentration, to focus on one thing and to do it well. Distractions in this modern society are just numerous. To concentrate and persist on resolving difficulties and unknowns is a valuable characteristic that will be extremely helpful all along my career.

The thesis is difficult because of obstacles and unknowns. Barriers of all kinds have lay before me one after another impeding me in advancing the subject. Fortunately, the real world truth is that things do go better when sufficient efforts are made thus believes are really important.

Lots of times people asked me why bother a PhD after my diploma of engineering with which I can already find a good job. My answer is that I am doing intelligent robotics, a domain still within its infant age while with great potential to be the next culmination of modern technology. I consider it beneficial to have a academic research training of the state-of-the-art robotic technology. Three years are not short, so it is better to take the chance now than later. If it becomes true one day that the robotic industry brings huge economic benefit, I will then ready to be a specialist in robotic software or the intelligent algorithms, and I prepare for that moment.

The next important thing should be to give my thanks. A lot of people have helped me in different types of ways during my life as a Ph.D candidate. First and foremost, I would like to give my sincere thanks and appreciations to my PhD advisor Mr. Stéphane Régnier for his insightful ideas and patient mentoring. I will certainly not be able to fulfil the thesis without his full support. His strict and rigorous attitude to academic research motivates me to overcome the most extreme difficulties encountered all along my days at ISIR. I would also give my thanks to Mr. Sinan Haliyo for his technical support on programming.

I should certainly appreciate also Mr. Ryad Benosman and Mr. Sio Hoi Ieng at Institut de la Vision for their judicious instructions and advices on numerous vision algorithms and their assistances on article writing. I still remember the several months that I have spent at vision institute making effort to develop the shape registration algorithm.

Next, great thanks should be given to Ms. Cécile Pacoret and Mrs. Aude Bolobion for their suggestions in the advancement of my thesis, their willingly collaborations on my work and their constructive and sometimes critical comments on my paper drafts. More thanks will also be given to Jean-Ochin Abrahamians and Allister Fourtet for their English proofreading.

I will certainly not forget Abdenbi, Bruno, Sebastian, Puco, Tiantian, Soukenya, Lora, Tianming, Nicolas and Mokrane in by bureau at ISIR for everyday chat and discussion, either professional, personal or recreational and as well Joao, Cédric, Charles, Xavier at Institut de la Vision.

Moreover, great appreciations should also be delivered to my parents and all my family in China, who have never ceased to support me morally all along these days.

Finally, if I happen to forget someone, I would like to use this final chance to thank all the anyone else that have assisted me in fulfilling this thesis ...

14/09/2013
In Paris

Contents

General Introduction

Capturing and analysing high speed movement of a flash of a second in every single detail, e.g. the breaking down of a fallen cup or a bullet hitting through a pear, is not a fancy story nowadays. Such videos can be easily found on Youtube since high speed cameras now enter many people's horizon. It is becoming an indispensable tool for photography professionals, journalists, engineers or researchers from different domains.

In robotics, a fundamental task is to *process* the acquired visual information for sensory motor control or for merely scene understanding. The acquisition being ultra fast, the processing speed inevitably becomes the bottleneck even when employing the most state-of-the-art technology. Computers strive to process the acquired images in real-time not more than 30 frames per second. The huge amount of data flow overwhelms the memory and the computation power of current CPUs. In many robots, it can often be observed that once the computer vision is turned on, the embedded systems start to suffer.

What will be a reasonable solution to overcome this difficulty? As you may be aware, the generation of images is only for the purpose of human visualization and actually produces no good when it comes to processing. There is no mandatory necessity for robots to acquire a complete image and afterwards extract the scene information from it. On the contrary, basic level processing should be performed at the same time as in visual acquisition, just as the retinas of animals or human beings do. The information is encoded in the form of electrical spikes (coined as "events" throughout the thesis) that are transferred asynchronously and continuously on nerves and between neurons. Why cannot our cameras and computers acquire and process the visual information in a similar way? As a milestone, Mahowald and Mead gave birth to a pioneer prototype of neurmorphic silicon retina in the year 1992. It opens way of a revolutionary visual processing paradigm, which no longer records light intensities, but electrical spikes that indicate spatial or temporal contrast changes, in the way all living beings do. In this thesis, we adopt this new bio-mimetic approach to process visual data, termed as *event based vision*. It will be shown that its encoding efficiency goes far beyond the conventional frame based approaches by orders.

The event based vision framework allows highly compressed visual data acquisition and high speed processing. Its internal asynchrony not only permits faster but also more robust vision algorithms. Following the completion of the first ready-to-use silicon retina, called Dynamic Vision Sensor (DVS) and its related drivers and utilities, it comes to an era when innovative and ambitious idea of mimicking living retinas can seriously be taken into real-world applications. This thesis is one of the first trials that employ an event based silicon retina in two real-world microrobotic applications.

The reason why the silicon retina is tried on microrobotics as one of the first

applications is not incidental. High speed visual processing can benefit a large variety of microrobotic applications. Since mass decreases by cube order while surface decreases by square order, surface forces distinguishably overweight the gravity force in the microworld. The movement is often instantaneous due to negligible masses and highly dynamic phenomenons occur frequently under the impact of numerous force fields, the mechanisms of most of which unfortunately still remain unclear.

The requirement in haptic feedback teleoperated micromanipulation arouses further interests in high speed vision. As the manipulation handle in human hand, a haptic device is firstly a tool to control the movement of objects. But unlike conventional joysticks, the device can meanwhile feedback forces to operators thus allows interactive sensation of the manipulation process. In order to achieve a realistic tactile sensation and to maintain a stable system, a 1 kHz sampling rate is required. As will be shown, vision is an indispensable tool for force feedback micromanipulation due to the complexity of mounting force sensors on end-effectors. The state-of-the-art vision algorithms that attain this high frequency are too simple to be robust. Therefore, an emergent technology demand to process visual data at more than 1 kHz in the domain of micromanipulation provides a perfect test bench for the experimentation of the novel event based visual framework.

Two concrete applications are thus developed. A popular method to investigate interactions and forces on mechanical properties or bio-medical functions of micro-object is to use optical tweezers. They are able to grasp minuscule objects by tightly focusing a laser beam. If the trapped object is a microsphere, the physical model indicates that the distance between the sphere and the laser center is proportional to the optical forces exerted on that sphere. Therefore, high speed position detection of microparticles becomes an essential problem. In chapter 2, the event based vision is used to track positions of the bead with known radii and feedback forces in two-dimensional space. A continuous Hough transform circle detection algorithm is developed for this purpose, which is able to run at 30 kHz. The experimental setup is based on an already mounted 2D optical tweezers system. In chapter 3, the opto-mechanical set-up is redesigned to enable simultaneous multiple traps in the whole three dimensional space. To detect spheres in 3D space, the previous tracking algorithm on fixed radius circle is no longer adapted. A robust circle fitting algorithm working on event data is thus developed to estimate both the position and the radius of a circle at high speed. As a result, the system turns out to be very successful and a patent has been submitted for this invention. It is then meant to be maturated into a commercialized product in the near future.

In current micromanipulation technologies, the manipulation end-effectors are most probably the canonical mechanical ones rather than non-contact methods as optical tweezers. In this case, a haptic device is not meant to feedback real interactive forces but simply to assist operators performing difficult tasks or blocking dangerous operations. Ensuring stability in such applications has always been an important issue that requires also a high sampling rate of 1 kHz. Chapter 4 resolves this problem by designing an event based shape registration algorithm. It uses a pre-constructed model to track an arbitrary non-deformable form at a speed of 4

kHz. In the experiment, we will show how to use this algorithm to stably feedback forces from mechanical micro-tools, such as a piezoelectric microgripper.

Chapter 1

High Speed Vision - From Frame Based To Event Based

Contents

High speed vision deals with the processing of visual data in a speed much higher than the usual frame rate. In this chapter, the methodology of visual processing will be shifted from the conventional frame based method to the novel neuromorphic event based method. Some of the fundamentals of the event based approach and the potential applications on robotics and especially on microrobotics will be introduced.

1.1 High Speed Cameras

Cameras are indispensable professional or recreational tools now permeating into many aspects of our daily life. *High speed* cameras, as the name suggests, are capable of capturing fast motions that cannot be perceived by human eyes. Expensive although, these devices arouse interests and curiosities of thousands of fans and photography professionals. The wonder of high speed cameras lies in its capability of capturing slight transient movement in a flash of a second far beyond the ability of human eyes. It becomes not only a toy for geeks or artists, but also a serious device of broad reach applications, like sport performance analysis, journalism, scientific, industrial and military research. As the Vision Research Inc. [Vis] states:

> "High speed photography is an engineering tool, much as is an oscilloscope or a computer."

1.1.1 Image Data Acquisition

With the fast development of modern electronic technology, commercialized industrial cameras are now able to record images at at least several hundred frames per second (fps), contrary to normal motion pictures or television, in which images are filmed and played back at only 25∼30 fps. High speed applications require an even higher frame rate in the order of kilo frames per second (kfps). High speed cameras thus come into play and gain its position. A well-known industrial example is the car crash examination. High dynamic transient status of a car and its passengers in the moment of crashing are recorded on a local flash disk of the camera and details are analysed thereafter frame by frame off-line. The same technique also plays an important role in sports training. Fig.1.1(a) shows a picture taken by a high speed camera when a bullet is piercing through a poker card and Fig.1.1(b) shows a bullet hitting an apple. Transient details are captured by high speed cameras by two orders of magnitude higher frame rate than the limit of human eyes. Nowadays, a middle-end high speed camera can usually acquire several kfps in full resolution (e.g. 1024×800) and several hundreds of kfps in highly reduced resolution (e.g. 64×8).

The research and commercialization of high speed cameras accelerate in recent years. An ultra fast camera designed at University of Rhode Island can run at a speed of 800,000 frames per second [Fas]. Researchers at Massachusetts Institute of Technology (MIT, Cambridge MA) have developed a variety of cameras that can be used to create a trillion frames per second – capturing even the motion of speed-of-light photons. Commercial high speed cameras use either a charge-coupled device (CCD) or a CMOS active pixel sensor as pixel component. Some of the manufacturers and resellers are: Basler, Dalsa, Matrox, Photron, Prosilica, Stemmer Imaging, Olympus, which are the major providers, and vision research Inc., AOS technologies AG, Fastec imaging and Optronis, which are smaller but more specialized ones. Fig.1.2 shows one of the industrial camera model from Basler AG.

(a) (b)

Figure 1.1: (a) A photo showing a bullet piercing through a poker card. All the details taken place in a split of a second are captured by the high speed camera when human eye fails. Image courtesy of verycoolpictures.com. (b) A photo of a bullet hitting through an apple. Image courtesy of athousandandone.com.

In spite of the wonders brought by these exciting devices, their drawbacks are significant due to the huge amount of image data generated. Image data is usually stored on a local memory of the camera and retrieved off-line thereafter. This does no harm to most of the applications but does cause problems when on-the-fly transmission and processing are required. Since the data quantity is huge, the duration of recording is thus limited to only several seconds. For example, the acquisition of uncompressed 1024×1024 three-channel high resolution color images at 3 kfps during 1 second will occupy 9 gigabytes space on the disk ($1024{\times}1024{\times}3{\times}3000fps = 9GB/s$). Therefore, the applicabilities of this type of sensor is restricted to those off-line studies of transient phenomena.

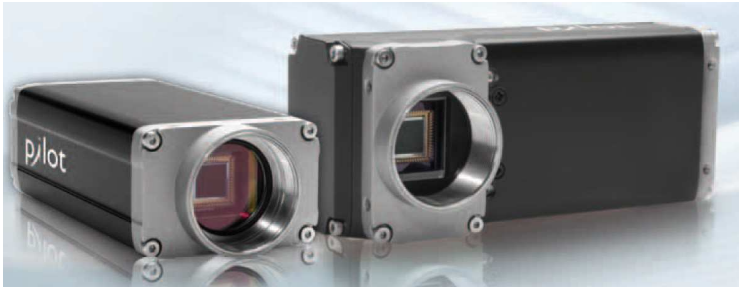


Figure 1.2: Basler high speed camera, Pilot piA640-210gm model, 648 x 488 resolution, 210 fps, monochrome, with GigE Vision interface. The camera achieves 210 fps at full resolution. Frame rate can still be increased by reducing the resolution. This is an industrial camera that can adapt to entry level high speed applications in reduced resolution mode with a moderate cost, typically several hundred euros. Image courtesy of Basler user manual Pilot series.

1.1.2 Image Data Transmission

Applications in robotics or industrial machine vision oblige low latency real-time image data transmission. The term "real-time" in this thesis refers to a system responding immediately after receiving the input signal and having latency short enough to be really perceived by users. The contrary of this term is "off-line". The definition of "real-time" in software engineering is not adopted, which is used to indicate the high precision of timing that can be controlled by software in an operating system.

For data transmission, several high speed data transfer interfaces are standardized for this purpose, which are GigE, Firewire, Camera Link and USB 2.0. Fig.1.3 shows some of these well-recognized trademarks. The different types of camera interfaces give machine vision customers the flexibility but also cause the inconvenience in understanding the advantages and limitations of each alternative.

Gigabit Ethernet (GigE), with its maximum data rate of 1 gigabit per second (Gbps) is a fast emerging machine vision interface. The GigE Vision camera's control registers are based on a command structure called "GenICam". This structure is built through the norm of the European Machine Vision Association (EMVA). The standardization of programming interfaces is essential for an easy use of fast cameras because different manufacturers can provide common software thus enhance hardware inter-operability and encourage community driven software development. The GigE interface is currently a strong and promising competitor to the older Firewire interface, because GigE standard supports longer cable length (100m), easier camera network protocol and stronger inter-operability, which are all important factors for industrial field automation and robotics.

Figure 1.3: The mostly used industrial high speed camera interfaces (from left to right): GigE vision, CameraLink and Firewire.

CameraLink has the fastest parallel communication rate among the mentioned four but requires the expense of an extra hardware - the frame grabber, to convert analogue video signals. It is thus the most expensive solution but offers the highest data throughput (over 5.4Gbps) and the lowest latency. The CPU load can also be reduced thanks to the independent hardware frame grabber. One disadvantage is that the distance between the camera and the host PC is limited to only 10 meters.

The firewire IEEE 1394 is designed with easy use in mind, since no external power supply is required and the port is equipped on most desktop PCs. However it is gradually becoming obsolete. Its disadvantage is that the data rate is low, which

	GigE vision	CameraLink	Firewire (IEEE 1394 a/b)
Data Rate	125MB/sec	680MB/sec	50-100MB/sec
Max Cable Length	100m	10m	5m
Frame Grabber	No	Yes	No
Cost	Low	High	Low
CPU Load	High	Low	High

Table 1.1: Comparison of different types of high speed camera interfaces (1 Byte = 8 bits). Note that theoretical data rate can hardly be achieved due to communication overhead.

theoretically achieves up to 480Mbps. Like other interfaces, theoretical rate can hardly be attained due to overhead of communication protocols.

USB 2.0 is a universal data interface and has not yet been standardized for high speed cameras. Since USB serves both as power supply and signal transmission, the power offered by a USB port may not be sufficient for numerous types of cameras. In Table 1.1, different types of high speed camera interfaces presented above are compared.

The data transmission rate provided by the above interfaces is high but in many cases still not sufficient. Let us consider a high speed camera that generates 1 mega pixel (1000×1000 resolution) grey level images at a rate of 1 $kfps$. Over 1 gigabytes uncompressed image data is acquired every second, which is beyond the capacity of the fastest CameraLink interface. To reduce the data quantity for real-time transmission, a general solution is to limit the imaging window, namely to specify a region of interest so that only a partial imaging area is active. Another possible way is to subsample high resolution images to a manageable level. With a moderate resolution (e.g. 350×290 pixels), a real-time transmission rate in the order of 6 kfps is the state-of-the-art. However, it will be shown next that the limit in data transmission is a, but not the only critical bottleneck in real-time high speed robotic applications.

1.1.3 Image Data Processing

Previous sections have shown that the modern technologies favour high speed image acquisition and is somehow limited by real-time data transmission. Once images have been transmitted to a host computer, however, a more serious bottleneck emerges – the processing.

Real-time high speed visual processing, which by definition requires a processing frame rate at minimally several hundred frames per second, is an almost unexploited area. The published ultra fast image processing algorithms in the order of kHz usually expose two principle limitations. Firstly, the original image is either subsampled or its field of view is reduced to a region of interest (ROI). The disadvantage of using ROI is its highly constrained working space, e.g. 80×80 pixels. Object may easily go out of sight during the tracking procedure. Secondly, the algorithms applied

are restricted to the most simple ones to avoid important CPU loads, sometimes hardware implementation is even compulsory. One of the most commonly used algorithms for high speed micro-particle tracking is the center of mass $(\overline{x}, \overline{y})$ (or called centroid), which can be calculated as follows:

$$\overline{x} = \frac{\sum_{x=0}^{w}\sum_{y=0}^{h} xf(x,y)}{\sum_{x=0}^{w}\sum_{y=0}^{h} f(x,y)} \quad \overline{y} = \frac{\sum_{x=0}^{w}\sum_{y=0}^{h} yf(x,y)}{\sum_{x=0}^{w}\sum_{y=0}^{h} f(x,y)} \tag{1.1}$$

where $f(x, y)$ is an arbitrary function that defines the value in the pixel location (x, y). w and h are the width and the height of the concerned region. This algorithm has been implemented in the commercialized software LabView.

Another frequently used algorithm is the image moments, which can be defined by:

$$M_{jk} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^j y^k f(x,y)dxdy \tag{1.2}$$

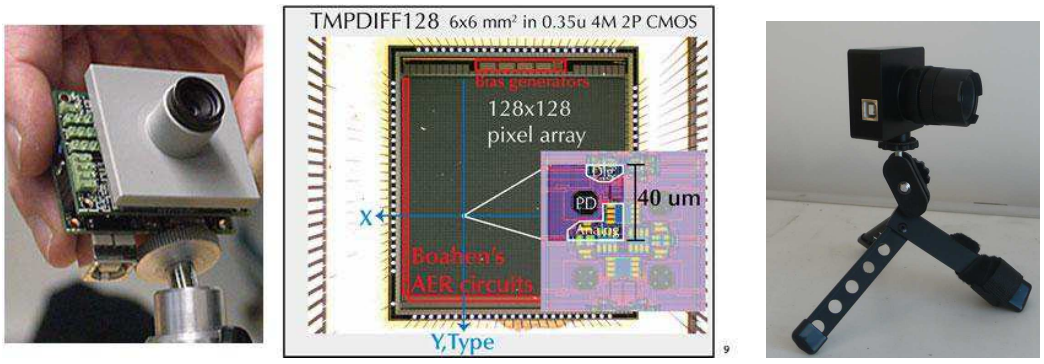
The moment of the function $f(x, y)$ can be used to define a group of shape features. j and k can theoretically take all non-negative integer values and define the order of the moment. Practically, zero to second order moments are calculated.

It can be clearly inferred that no complex tasks can be performed using such simple vision algorithms. A micro-organism tracking system with a resolution of 128×128 running at 1 kHz has been reported, which uses a hardware implementation. As expected, the dedicated hardware simply calculates the centre of mass and moments [Oku 2005], as Equation 1.1 and 1.2 show. Unluckily, lots of well-established, robust and complex vision algorithms out there are all computational expensive. Therefore, high speed vision in the order of kilohertz in a large working space has always been one of the most challenging and seemingly unsolvable tasks until a revolutionary bio-inspired solution is found.

1.2 Silicon Retinas

1.2.1 Neuromorphic Engineering

The solution we propose comes from neuromorphic engineering. This is a newly developed research area, of which the concept were first established by Carver Mead in the late 1980s [Mead 1989] at CalTech, USA. This discipline uses VLSI (Very Large Scale Integrated circuits) and software to implement models of neural systems for perception and sensory-motor processing. For years, the morphology of neurons inspired scientists to build circuits and systems that automatically execute complex computations that mimic the behaviours of the biological world. This interdisciplinary research domain takes results from biology, mathematics, computer science and electronic engineering to design artificial bio-mimetic systems. A typical example of such neuromorphic devices that will be used throughout the thesis is the silicon retina which mimics human eyes. Fig. 1.4(a) shows a photo and the circuit layout of such retinas.



(a) (b)

Figure 1.4: (a) (left) A silicon retina prototype (ETHZ, Switzerland) based on the Address Event Representation (AER) and (right) its circuit layout. (b) A photo of Dynamic Vision Sensor (DVS), a first practically usable silicon retina prototype from ETHZ.

A distinctive feature of biological neural systems is their spiking mechanism. Spikes only exist between neurons. The spikes or "events" (term exploited in the forthcoming chapters) communicate asynchronously. The Address Event Representation (AER) is such a communication protocol for transferring these events between neuromorphic chips. On an AER based system-on-chip, a unit (a neuron) triggers a transmission when its internal state surpasses a threshold. The change of the internal state is originated either from the outside stimuli or the inputs from other units. Each unit is independent and triggers the transmission request independently. The data transferred contains the address of the spiking unit (spike source) and its polarity (positive or negative). Similar to the neural system, the frequency of the spikes encodes the intensity of the outside stimuli. The communication bandwidth is thus unequally occupied by the busiest spiking unit. This results in an activity driven system that lacks of redundancy, runs at high frequency and produces small amounts of data, similar to what happens within the biological neural systems.

1.2.2 Dynamic Vision Sensor (DVS)

It is a fact that biological vision systems outperform human engineered vision devices in almost every aspect. Motivated by nature's superiority, the neuromorphic community is developing novel types of vision devices that - like their biological models - are driven and controlled by "events" happening within the scene in view. Whereas conventional image sensors are driven by artificially created synchronous timing and global control signals that have no relation to the source of the visual information. In the following, the conventional image based processing with a fixed sampling interval will be termed as "Frame Based" and our asynchronous neuromorphic processing will be coined "Event Based".

Over the past few years, a variety of these "event-based" vision sensors have been

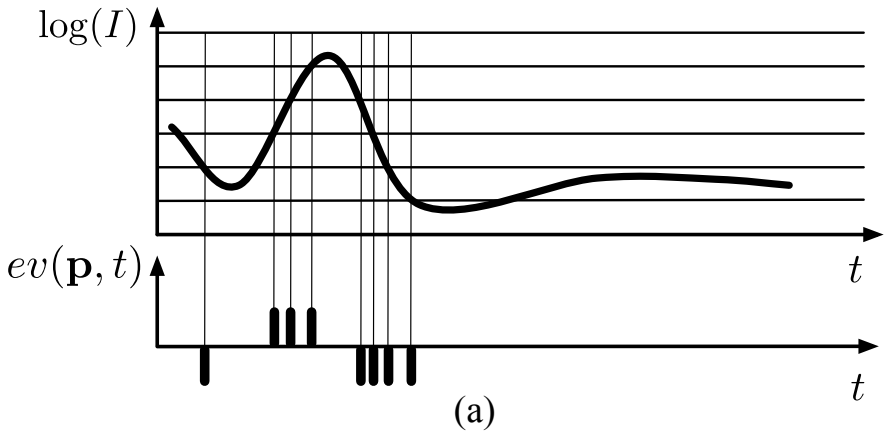


Figure 1.5: Principle of the generation of events of DVS's pixels, adapted from [Lichtsteiner 2008]. Events with +1 or -1 polarity are emitted when the change in log intensity exceeds a predefined threshold.

implemented, including: temporal contrast vision sensors that are sensitive to relative light intensity change, gradient-based sensors that are sensitive to static edges, edge-orientation sensitive devices and optical-flow sensors. Most of these vision sensors output visual information about the scene in the form of asynchronous address events (AER) [Boahen 2000] as other neuromorphic devices, thus greatly improving temporal resolution and increasing dynamic range as compared to conventional imagers. In most cases, the visual information is encoded in the time dimension and not as voltage, charge nor current. The pioneering work has to mention Mahowald [Mahowald 1992]. An overview of recent developments in the field can be found in [Delbruck 2010].

The Dynamic Vision Sensor (DVS), see Fig.1.4(b), is an AER silicon retina prototype with 128 × 128 pixels. It is the first practically usable prototype developed in ETHZ, Switzerland. Most other neuromorphic labs focus heavily on proof-of-concept hardware development thus the device is difficult to be used by end users. DVS is a USB enabled camera with the ability to adjust online the camera parameters. The use of this sensor on robotic applications is possible without the knowledge of the inherent hardware structure.

The DVS output consists of asynchronous address events that signal scene reflectance changes at the times they occur. Effective visual data compression rate directly depends on the dynamic contents of the scene. Each pixel detects changes in log intensity larger than a threshold since the last emitted event (typically 15% contrast) to increase the dynamic range. When the change in log intensity exceeds a set threshold, an ON (+1) or OFF (-1) event is generated by the pixel depending on whether the log intensity has increased or decreased. The basic principle of DVS's response to temporal contrast is illustrated in Fig.1.5. The redundancy free encoding of the visual information from a dynamic scene, delivered by this camera, is naturally suited for real-time processing in high-speed vision tasks like object

recognition and tracking. The advantages of using this type of sensor in visual computation will be presented in detail in Section 1.3.

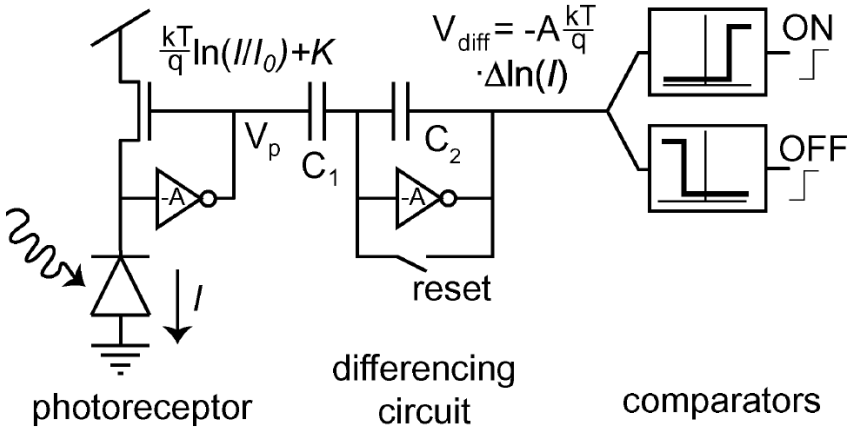


Figure 1.6: Abstracted pixel circuit schema, image courtesy of [Lichtsteiner 2008]

Fig.1.6 illustrates an abstracted pixel circuit schema of a single DVS pixel. The pixel level circuit is composed of three parts: the photoreceptor, the differencing circuit and the comparators. The photoreceptor circuit automatically controls the gain of an individual pixel by a logarithmic response while responding quickly to illumination's changes. The differencing circuit is used to remove the DC current and reset the initial level after the generation of an event. Combined with the logarithmic conversion in the photoreceptor, the pixel is thus only sensitive to temporal contrasts. The comparators are then used to thresholding the voltage to generate "ON" or "OFF" events. Under this design, only the changes of light intensity will generate data. Since the retina uses the AER protocol, the transferred data not only contains the polarity (+1 or -1) of the change, but also the pixel location x, y and the timestamp t of that change. A complete description of the technique specification and of the variant properties of DVS are beyond the scope of this thesis. Interested readers are recommended to refer to the seminal paper [Lichtsteiner 2008].

1.2.3 Asynchronous Time Based Image Sensor (ATIS)

The event-driven time-encoding sensors, especially vision sensors, have never ceased to evolve. They have matured to such an extent that they appear to offer not only a viable, but in many applications, a superior alternative to conventional image sensors. The latest generation of event-based, time-domain encoding image sensors - ATIS ("Asynchronous Time-based Image Sensor") [Posch 2008][Posch 2011] is the most recent prototype of silicon retinas. This sensor contains an array of fully autonomous pixels that combine an illuminance change detector circuit as that in DVS and an *additional* conditional exposure measurement device. The change detector individually and asynchronously initiates the measurement of an exposure/grayscale value only if - and immediately after - a brightness change of a certain magni-

tude has been detected in the field-of-view of the respective pixel. The exposure measurement circuit in each pixel individually encodes the absolute instantaneous pixel illuminance into the timing of asynchronous spike pulses, more precisely into inter-spike intervals (Fig. 1.7). This principle is also referred to as asynchronous pulse-width-modulation (PWM) imaging [Chen 2011].

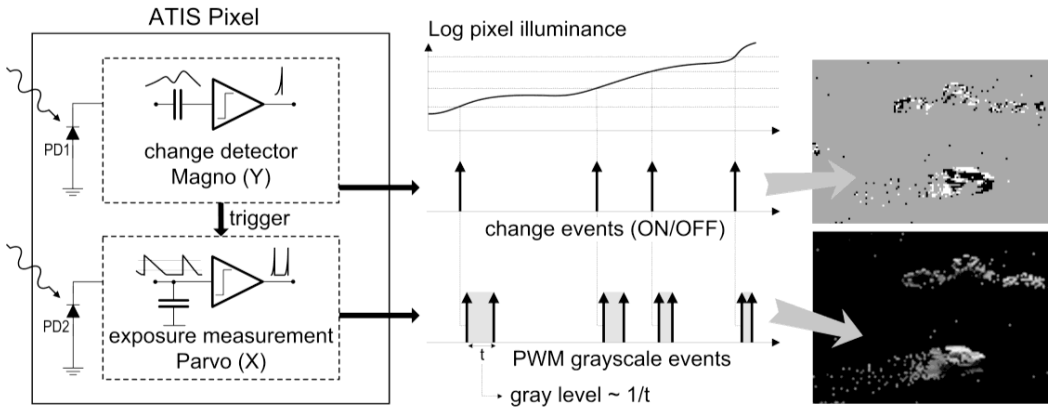


Figure 1.7: Functional diagram of an ATIS pixel [Posch 2010]. Two types of asynchronous events, encoding change and brightness information, are generated and transmitted individually by each pixel in the imaging array. Grey level is integrated immediately after a binary polarized event is generated.

Like DVS, ATIS also uses the AER protocol for its circuit design. Pixels do not rely on external timing signals (shutter, frame clock) but autonomously request access to an asynchronous and arbitrated output channel when they have relevant information, i.e. new grey scale values (like the polarities in DVS), to communicate. At the readout periphery, the asynchronous event-encoded grey level is arbitrated, furnished with the pixel's array address by an address encoder and sent out to an asynchronous bit-parallel AER bus. Fig.1.7 shows a functional diagram of the ATIS pixel. The time-domain encoding of the intensity information automatically optimizes the exposure time separately for each pixel instead of imposing a fixed integration time for the entire array, resulting in an exceptionally high dynamic range and improved signal-to-noise-ratio. Similar to DVS and all event based sensors, the pixel-individual change detector driven operation yields almost ideal temporal redundancy suppression, resulting in a maximally sparse encoding of image data.

In this thesis, DVS is used due to its matured drivers. The purpose in the first step is to validate the approach of employing event based sensor and algorithms for high speed object tracking in microrobotic applications. However, DVS exposes several drawbacks, which are the lack of static edges and of low resolution. Therefore, ATIS is actively being developed in the meantime to include gray levels and to achieve higher resolution. Replacement of DVS by ATIS is envisioned in the near future once its drivers and user interfaces are ready, which can instantly improve

the quality of the microscopic object tracking.

1.3 The Advantages of Asynchronous Event-based Vision

1.3.1 Frame Based Methodology

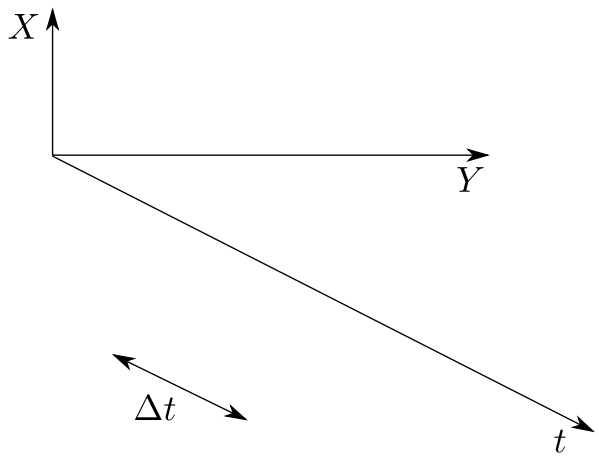


Figure 1.8: Frame based encoding of visual information. Production of unnecessary repetitive background data in successive frames leads to high computational load and memory consumption.

To understand the advantages of event based computation, the limitations of the frame based acquisition and processing should be firstly comprehended. Let $I(t)$ be an image acquired at time t, so an image sequence can be written as $I_{seq} = \{I(t_0), I(t_0 + \Delta t) \ldots, I(t_0 + n\Delta t)\}$. Δt is the fixed sampling interval between frames. A large amount of background data recorded in successive frames of I_{seq} is purely repetitive, see Fig.1.8. The higher the acquisition frequency, the more redundant information the camera generates and waits to be processed. The production of unnecessary redundant data leads to high computational load and memory consumption. Furthermore, the frame based processing algorithms rely on the iteration of all the image pixels several times and then repeat it for each frame.

Circular object tracking is a common task in computer vision and is a decent example to illustrate the inconvenience of the frame based method. First of all, a Gaussian filter should be applied to smooth the image and to reduce noise. Secondly, an edge detector, e.g. Canny edge, is used to extract the contour of the objects, both foreground and background. Finally, a Hough transform [Hough 1959] may be chosen to detect multiple circles from the edges extracted during the previous procedure. As every frame based methodologies, most of the procedures above require looping over every pixel in the image; the algorithm complexity increases linearly with the pixel number. Hence artificial frame based paradigm stays away

from the way biological beings treating the visual information in the retina and the visual cortex.

1.3.2 Event Based Acquisition

The neuromorphic silicon retina provides a revolutionary way for visual acquisition. It changes the way of how visual data is acquired, transmitted, encoded and processed. In the following text, DVS is treated as a representative of silicon retinas. The advantages of event based acquisition over conventional frame based imagers lie in several aspects.

Event based acquisition is *redundancy free*. A silicon retina outputs compressed digital data in the form of events responding to contrast changes. Only the dynamic information stimulated either by a moving object or by the change of light intensity, e.g. defocus, is recorded.

Event based acquisition is *asynchronous*, meaning that each pixel can trigger an event independently from others once light intensity change is perceived. No global clock is used, unlike the conventional frame based cameras that are sampled by a fixed interval. The timing of the DVS events can be conveyed with a minimum temporal resolution of 1 µs. Thus, the "effective frame rate" is typically of several kHz. The DVS's pixel bandwidth attains to 3 kHz under 1 *klux* scene illumination. Low illumination conditions will reduce the pixel bandwidth, which may be the case in micromanipulation.

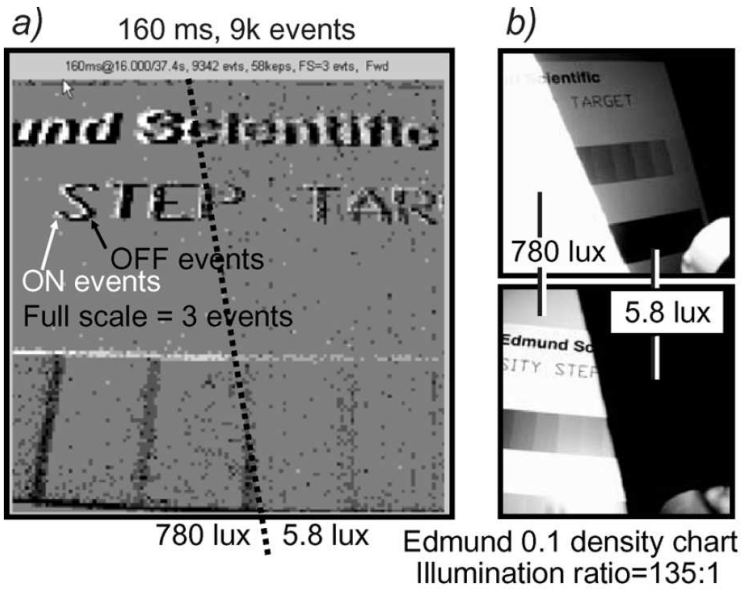


Figure 1.9: Illustration of DVS's high dynamic range property. (a) Output from DVS viewing an Edmund density step chart with illumination ratio of 135:1. (b) The same scene captured by a conventional frame based digital camera to expose the two halves of the scene. Image courtesy of [Lichtsteiner 2008].

Pixel size (µm)	40×40	Resolution	128×128
Timestamps precision	1 µs	Pixel bandwidth	3 kHz
Response latency	15 µs	Dynamic range	120dB
Power consumption	2.3mW @ 3.3V		

Table 1.2: The main properties of Dynamic Vision Sensor.

Event based acquisition has *large dynamic range* (DR). In image acquisition, DR indicates the capability to capture light intensity without under- or over-saturation. DVS has dynamic range of 120dB. A large dynamic range is essential for autonomous robots working in different environments both indoors and outdoors. Fig.1.9 shows the DVS's high dynamic range capability to view scenes with an extreme intensity contrast. Where conventional cameras encounter over- or under-exposure problems, DVS reacts correctly with the light intensity of the scene, untouched by its extreme light conditions.

Other advantages of event based acquisition includes *low latency* (15 µs) thanks to the AER protocol. Low latency is especially suitable for robotic applications that require high speed sensory motor coupling. DVS achieves also a very *low power consumption* of only 23 mW. Table 1.2 summarizes the major advantageous features of DVS:

1.3.3 Event Based Processing

Event based processing allows a *high detection frequency*. The event flow intrinsically encodes objects' movements. This is why real-time high speed object tracking is the first task accomplished by DVS. The update of object's positions can be launched once an event is acquired, without waiting for the arrival of the entire frame, hence significantly increasing the system responsiveness to the outside stimuli. The consequent tracking speed can thus be one or two orders of magnitude higher than the conventional frame based correspondence. Fig.1.10 (a)-(h) shows event accumulation maps (AC map) of a character "H". An AC map is an easy visualisation of the accumulated events within a certain period of time regardless of their precise timings. It will be shown later that event based processing is capable of tracking the character "H" rotating at a very high speed of 1360 revolutions per second.

Event based processing permits *time based algorithms*. The timing of events is an extra essential element in visual computation. Frame based image processing algorithms substantially use the spatial correlation between pixels while ignoring the potential temporal relationship. Under the event based framework, each pixel independently spikes an event that is attached by a specific timestamp signifying the moment of the change. The precise timing brings critical information that could be extremely beneficial in algorithm design. Fig.1.10 (i) shows a 3D representation of events generated by a rotating "H". Emphasis will be drawn on the "third" axis representing the timestamps, which is an absent element in frame based processing.

Event based processing enables *more robust detection*. DVS intrinsically elim-

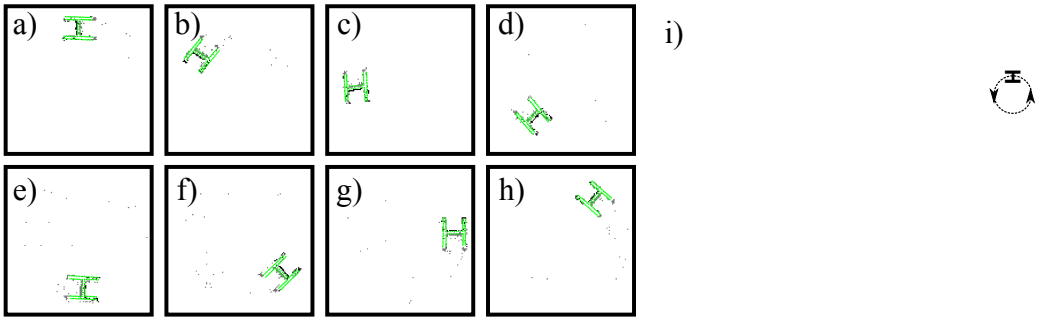


Figure 1.10: (a)-(h) shows event accumulation maps, namely accumulating events within a certain period regardless of their timings, of a character "H" rotating at 1360 revolutions per second. (i) is a 3D representation of the generated events from the rotating "H" with the vertical axis as time.

inates the background and the static objects thus reducing the need of an image segmentation, which is usually a long and error prone process in frame based processing. The time based algorithms are capable of considering the spatial and temporal correlations as a whole entity thus usually achieving better results when multiple objects collide. Moreover, the ultra-wide intrascene dynamic range (120dB) allows for robust operations under uncontrolled lighting conditions for outdoor mobile robots.

Finally, event based processing brings *zero energy consumption for static scenes*. As portable digital devices become more and more widely distributed, power consumption tends to be one of the upmost important issue in consumer electronics. Zero static power consumption is an attractive advantage for consumer electronics applications. [Lee 2012] demonstrates recently a real-time human gesture recognition interface developed by Sumsung employing DVS, aiming to compete with the Microsoft Kinect®.

JAER (Java tools for Address Event Representation processing) [1] is an user-friendly open source software to control various types of neuromorphic event-driven devices and provides a large amount of event based processing methods, most of which are still under testing. The control interface to adjust DVS's parameters such as the pixel bandwidth, the event generation threshold and the "ON-OFF" polarity balance are provided to facilitate third party development.

1.4 The Fundamentals of Event-based Computation

Asynchronous event based algorithms and their applications are the main subjects to be addressed in this thesis. In this section, a general framework of event based processing will be formulated mathematically. Several basic notions and notations used throughout the thesis will be presented.

[1] http://sourceforge.net/apps/trac/jaer/wiki

A classical grey-scale image is a $M \times N$ array of positive values that measures the light intensity in each pixel of a CCD or a CMOS sensor. An image sequence I_{seq} can be written as a set of one-dimensional functions $f_{x,y}$ expressing the variations of each pixel's intensity as follows:

$$I_{seq} = \{f_{0,0}(t), f_{0,1}(t), \ldots, f_{M-1,N-1}(t) \mid t \in [t_0, t_0 + n\Delta t]\} \tag{1.3}$$

In the classic scheme shown by Equation(1.3), pixel's information is sampled at fixed intervals Δt. This type of representation is incompatible with a compact representation of data due to its unnecessary storage of redundant data. To overcome this problem, event based solutions get rid of the constant parts of $f_{x,y}$. That is to say, wherever there is no significant luminance change, $f'(t) \simeq 0$, the sampled data are simply discarded and only those values that are distinctly changed through time are kept. If the luminance changes $f'_{x,y}(t)$ gets above a predefined threshold, the signal is sampled and an event is generated. Thus, an event from the retina that occurs at time t at the spatial location $\mathbf{p} = (x, y)$ can be mathematically defined by:

$$e(\mathbf{p}, t) = \begin{cases} sign(f'_{x,y}(t)) & if\, |f'_{x,y}(t)| > \Delta I \\ 0 & if\, |f'_{x,y}(t)| < \Delta I \end{cases} \tag{1.4}$$

where ΔI is the minimum sensible threshold. As it can be seen, the values of $e(\mathbf{p}, t)$ can be respectively set to $+1$ or -1, which represent a positive or a negative change of contrast.

An event $e(\mathbf{p}, t)$ describes an activity in the spatio-temporal space. Similarly to biological neurons, the influence of an event will endure a certain amount of time once it has been spiked. This temporal property of events can be introduced in the form of a "decay function" applied to model this phenomenon. Let $S(t)$ be a spatio-temporal set of events active at time t, then one has:

$$S(t) = \left\{ e(\mathbf{p}, t_i) | e^{\frac{t-t_i}{\tau}} > \delta \right\}. \tag{1.5}$$

with τ being the time constant parameter of the decay function, δ a predefined threshold and t_i the event's generation time. This function signifies that an event will lose its weight exponentially as time passes.

For practical implementation, a simple way is to assume a decay function with binary weights. An event is supposed to have only two possible status, active or inactive, instead of the continuous exponential values as in Equation(1.5). In this way, all the active events contained in $S(t)$ can be stored in an activity buffer maintaining the events' generation order. If the event's timestamp indicates an activity duration longer than a predefined threshold $\Delta t = \tau \ln(\delta)$, it is eliminated from the activity buffer. This mechanism has a built-in data structure support in most programming languages known as First In First Out (FIFO). However, a FIFO introduces a filter effect because the decay time Δt has a direct impact on the bandwidth of the vision system, so it is an important parameter to be chosen carefully. The decay time can vary from 1 µs to $50ms$ according to the dynamics of the scene and the application requirements. 1 ms is a good choice in most cases.

If binary weights are used (FIFO implementation), the active events can be defined in a simpler form. One can use S_{t_1,t_2} to represent the spatio-temporal event set within the interval $[t_1, t_2]$ as:

$$S_{t_1,t_2} = \{e(\mathbf{p}, t) \mid t \in [t_1, t_2]\} \tag{1.6}$$

where $t_2 - t_1$ is equal to the decay time τ.

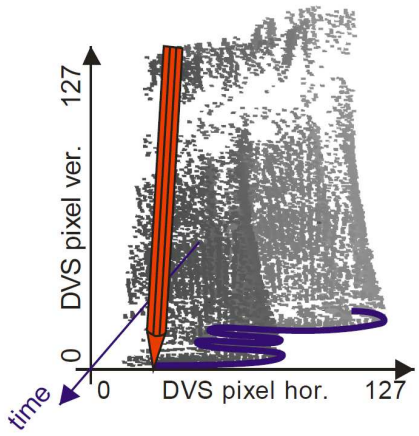
As illustrated in the definition, time is always a critical element in event based computation. In neuroscience, the "rate coding" or the "temporal coding" of spike trains are usually analysed. In "rate coding", we infer that most of the information about the stimuli is contained in the spike firing rate of the neuron, thus the frequency of spikes encodes the intensity of the outside stimuli. This is a simplified but widely used model in neural coding due to the large variances of inter spike intervals (ISIs) between successive spikes. In practice, the precise timing of ISIs may be hard to decrypt while the statistical properties are discriminating. In event based visual processing, the "rate coding" can be obtained by calculating the average event generation rate of the retina.

In other cases, "temporal coding" may be studied, in which the detailed characteristics of inter spike intervals are analysed and patterned to obtain more subtle and profound information. Readers interested in neural coding can refer to [Kandel 1991] for more explanations.

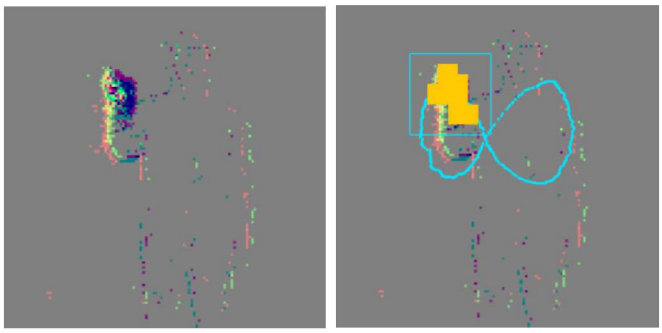
1.5 State-of-the-art of Silicon Retina Applications

The event based sensors are extremely adaptable to object tracking because of its spatial temporal encoding of movement. However, few successful applications have been demonstrated and few suitable algorithms are available due to its novelty. An earliest serious application has been developed and patented at the Austria Institute of Technology (AIT). An event cluster detection algorithm is used for vehicle detection in traffic surveillance. The algorithm is capable of grouping spatially adjacent events generated by vehicles. The group may change its size during tracking when vehicles approach, see [Litzenberger 2006a][Litzenberger 2006b]. This invention has then been commercialized as part of the Intelligent Transportation System (ITS) to count the number of vehicles passing by and to estimate their speeds.

Most of the state-of-the-art researches are conducted at Neuroinformatics Institute of ETHZ, Switzerland, where DVS was invented. A few event filters (background noise filter, direction filter, etc.) and a simple point cloud tracking, similar to what was used at the AIT, are presented in [Delbruck 2008], which is the first paper concentrating on frame-free event based algorithms. A ping-pong player has been built to demonstrate the sensor's high dynamic characteristics and the potential of coupling DVS with high speed sensory motor systems [Delbruck 2007]. In this work, the robot arm hits the balls immediately after they have arrived, in which a small latency of 2.8 ± 0.5ms is measured. In another work, a rod balancing robot is developed for the same demonstration purpose. The robot is used to stabilize a pencil



(a)



(b)

Figure 1.11: (a) Events generated by a pencil balancing robot to demonstrate the advantage of using DVS on high speed sensory motor applications. Image courtesy of [Conradt 2009]. (b) Real-time gesture recognition interface by stereoscopic DVSs, image courtesy of [Lee 2012].

by applying event based Hough line transform [Conradt 2009], see Fig.1.11(a). The 3D locations of the pencil are calculated by using a pair of DVS sensors mounted on different angles. The lightweight tracking and control algorithms are computed on an ARM7 microprocessor, which shows the practicability and the advantages of integrating DVS into embedded platforms with limited resources. Another application is the development of an object fall detector by tracking vertically falling point clouds, dedicated to elderly people home caring [Fu 2008]. Recently, a co-project between ETHZ and Samsung concerns a human gesture recognition interface based on a stereo pair of DVSs. It exploits more complicated event processing techniques. The motion trajectory of a moving hand is detected by spatiotemporally correlating the output events of the DVSs, see Fig.1.11(b). A neuromorphic approach is used to track correlated events and the stereo vision is used to strengthen robustness. Finally, gesture patterns were classified by applying a hidden Markov model (HMM) based method [Lee 2012].

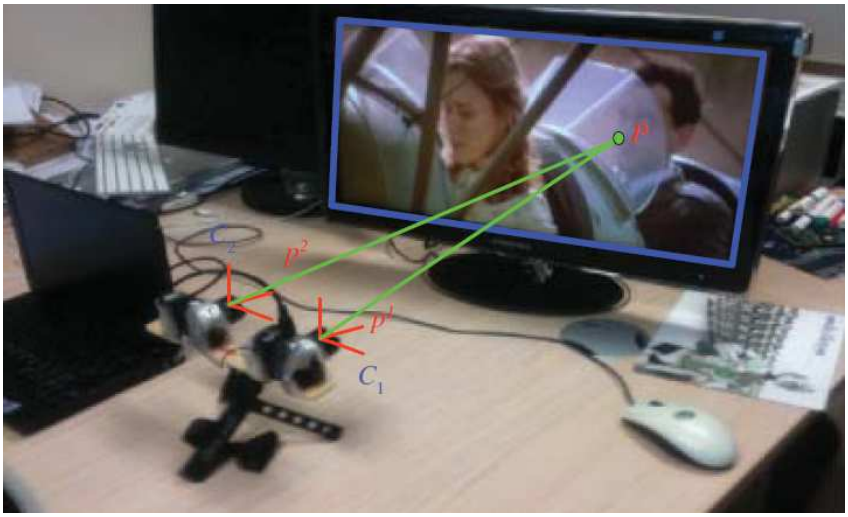


Figure 1.12: Two DVS retinas are placed in front of a liquid crystal display (LCD) screen. For each position of the screen, the probability coactivation field using the pixel activities is computed. Image courtesy of [Benosman 2011].

France participates actively in the most recent research in event based processing. Theories and new results on event based epipolar geometry [Benosman 2011] and stereo matching [Rogister 2011] have recently been published. In stereo-vision, one of the major problems is to find the corresponding points between binocular images. Conventional epipolar geometry exploits only the spatial relationship, while under the event based framework, precise timing of events becomes a supplementary yet essential element. Correspondences between binocular images can be verified for a second time by calculating the temporal difference between features. Fig.1.12 illustrates an experimental setup of the stereo-system using two DVSs. The method calculates the probability coactivation field using pixel activities. Another important advancement worth mentioning is the event based optical flow. Fig 1.13 shows the event based optical flow of a bouncing ball [Benosman 2012], of which the computation is significantly speeded up.

As can be seen, the use of DVS data is essentially different from the frame images that possess a fixed global sampling rate. The image processing community is accustomed to process images by looping over the entire image pixels with one frame as a basic processing unit. The major challenge of adapting conventional vision algorithms for DVS is the way of thinking. In our case, an event should be the basic processing unit and timestamp can be a useful computational element. Designing algorithms that respond to asynchronous, sparse yet accurately timed information is still not typical in conventional computer vision. Therefore, the new way of algorithm design consists of the major contribution of this thesis.

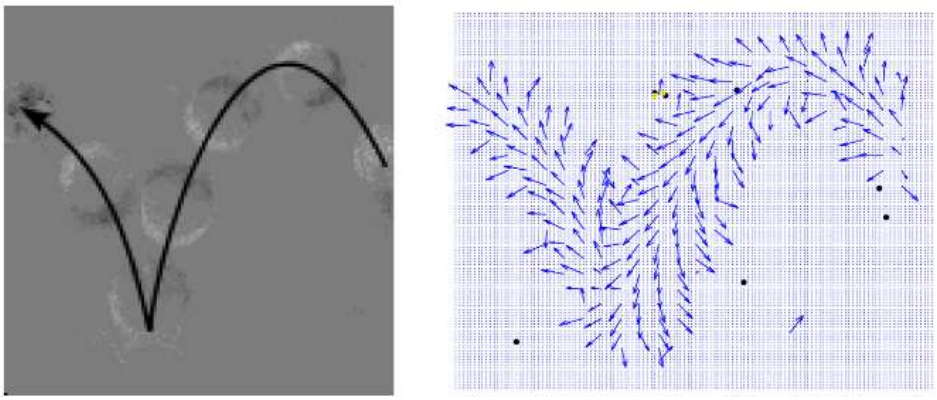


Figure 1.13: Optical flow of a bouncing ball: (left) decomposition of the movement, (right) the corresponding optical flow calculated by event based processing. Image courtesy of [Benosman 2012].

1.6 High Speed Vision in Robotics

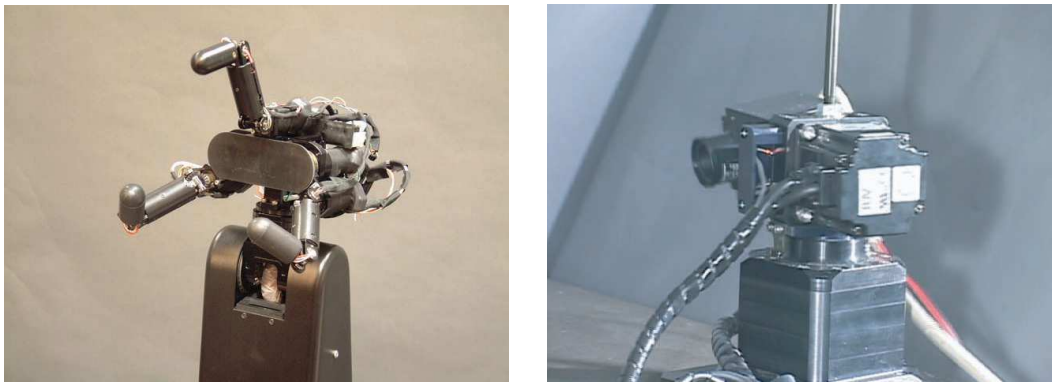
In this section, the reasons why the development of revolutionary high speed vision sensors and algorithms is critical in robotic applications will be briefly explained. The state-of-the-art and the underlying difficulties will also be briefly introduced.

1.6.1 Examples

Real-world robotic applications are real-time, a lot of which require a fast sensory motor system to fulfil dedicated tasks as quickly as possible. Numerous industrial robots [Lange 1998], visual product quality inspection [Noordam 2000], object shape measurement [Watanabe 2007], robotic manipulators [Namiki 2003] and mobile robots [Uchikura 2006] depend on high speed visual feedback to guarantee the productivity and the reactivity of the robots. The vision algorithms applied are only able to react swiftly in working environments with perfect experimental conditions. A throughput of the order of a thousand frames per second and a maximal latency of the order of milliseconds are usually required for high speed robotic control. In brief, the necessity of general purpose high speed vision solutions, including sensors and algorithms, is huge.

Two examples will then be introduced. An academic work consists in the development of a dexterous 3-finger hand at the University of Tokyo, see Fig.1.14(a). The 3-finger hand can close its joints at 180 degrees per 0.1s. The hand system is controlled by a massively parallel vision system, which is equipped with 128×128 photo detectors with an all pixel parallel processing array running at a rate of 1 kHz, see Fig.1.14(b).

A well-known industrial example is the product quality inspection, e.g. flat panel displays, electronic boards and postal parcels. Automated optical inspection (AOI) is the most effective approach that requires a fast frame rate to optimize the speed



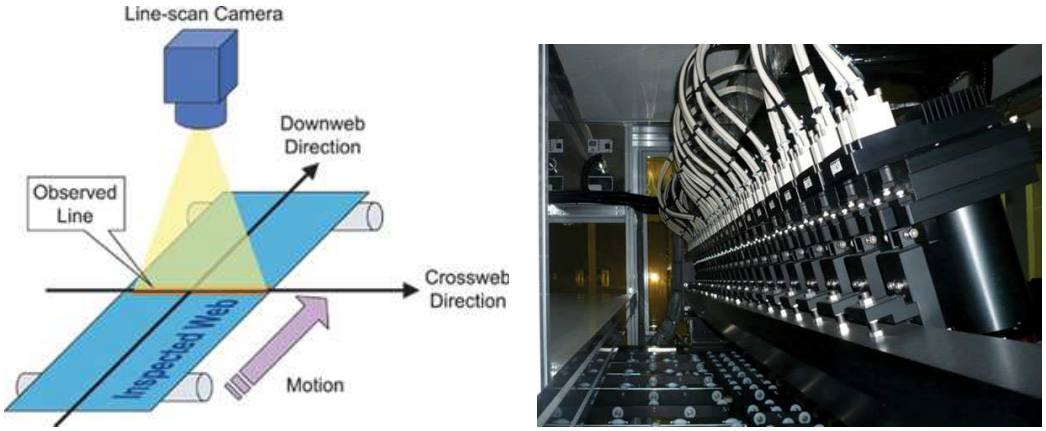
(a) (b)

Figure 1.14: (a) A dexterous 3-finger hand is developed at the University of Tokyo. (b) The hand system is controlled by a massively parallel vision system. It is equipped with 128×128 photo detectors with an all pixel parallel processing array. Image courtesy of [Uchikura 2006].

of inspection. AOI is required in flat-panel display production partly because of the industry's zero-defect policy. In these applications, line scan camera is almost the practically definitive solution. Line scan cameras, as the name suggests, record only one line of pixels at a time instead of a whole region as the commonly used area scan cameras. Despite the single dimension recorded by the camera, two dimensional images can be provided. The second dimension is obtained from the perpendicular motion of the observed scene relative to the camera scan line. In Fig.1.15(a), the camera is fixed while the scene moves with a known speed in order to reconstruct the 2D image. This type of camera can achieve a line scan rate of $10 \sim 90$kHz. Fig.1.15(b) shows a AOI network to inspect the defaults of plat panels. However, the use of line scan cameras is limited to the surface inspection thus not suitable to general robotic applications.

1.6.2 Difficulties

In the state-of-the-art vision based robots, massive parallel processing on hardware and GPU (Graphic Processing Unit) are seemingly the only solutions to overcome the speed bottleneck in conventional software based image processing [Dahmouche 2010]. Because of its different design principle, current parallel image processing algorithms are limited to: convolution-like filters for noise reduction [Chang 2006], center of mass (or Centroid) [Constandinou 2006], color based segmentation [Noordam 2000] and image moments [Namiki 2003]. The calculation of centroid or moments is illustrated in Equation 1.1 and 1.2. These algorithms require only the involvement of neighbouring pixels using basic addition or multiplication operations, thus are easy to parallelize. However, most of the vision algorithms are designed in a sequential way of thinking so a software implementation would be



(a) (b)

Figure 1.15: (a)The working mechanism of line scan cameras. The camera scans only one line at a time. The moving platform brings the object to be scanned forward. A 2D image can be reconstructed. The word "web" designates a stream of material, such as glass or paper. Image courtesy of ADR technology. (b) An AOI network locates and identifies repairable defects. If so, the panel is rerouted for possible repair. Image courtesy of Teledyne DALSA.

more understandable and cost-effective. The constraint of parallelism makes most of the existing middle and high level vision impossible to implement. Furthermore, the hardware development cycle is much longer than that of software given an identical task, because of the difficulties lying in expressing complex sequential ideas in a hardware description language, e.g. VHDL. Last but not least, the hardware system is not as flexible as its software correspondence. Off-the-shelf chips cannot be directly employed to implement vision algorithms. Customized chips are hard to be reprogrammed to perform different vision tasks. This inflexibility is especially undesirable during the integration of vision solutions to robots. In short, hardware is not the salvation.

To guarantee autonomy, mobile and humanoid robots require the use of embedded systems. Embedded systems generally include two categories. The first one includes microprocessors, microcontrollers and DSPs (Digital Signal Processor), which still require software design but on a smaller scale. The second category includes FPGAs (Field Programmable Gate Array) and alike. Their design is based on hardware description languages [2]. On one hand, embedded microprocessors are in most cases less powerful than a desktop PC. On the other hand, vision is always one of the most time consuming tasks in a robotic system. An embedded computation unit with low cost, low power consumption while still having real-time middle level processing ability, seems to be impossible. This tight technological constraint

[2]It is possible to configure a customized CPU on an FPGA platform and then program the CPU with a traditional programming language such as C. For example a Nios® II core can be configured on the FPGAs of Altera Corporation.

hinders the popularity of vision on such robotic systems.

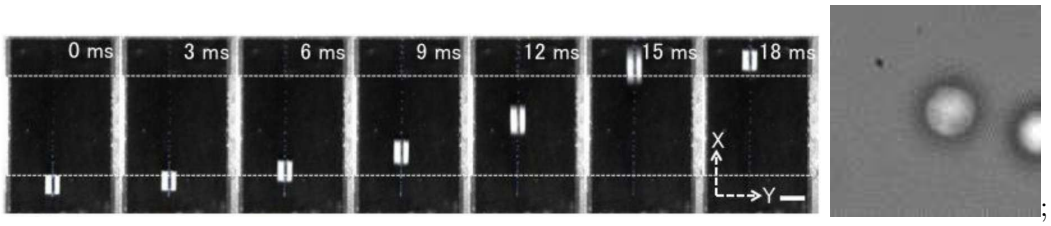
Researchers have then directed their attention to efficient biological visual systems. The event based vision sensors and algorithms are inspired by the spike mechanism from human retina and visual cortex. A more efficient encoding of the visual information brings a revolution in such a way that no grey scale data is recorded nor repeated. Less redundant information, less power consumption, faster processing ability and lower development cost are all-in-one achieved as will be shown in later chapters. Even if event based vision is still an emerging research area, real commercial success is foreseeable in a near future, to name a few, wireless sensor networks, stereo vision avatars, industrial manufacturing or inspection, human machine interface devices (e.g., eye-trackers, human body trackers), autonomous navigation systems, humanoid robots (e.g. iCub's eye) and visual prosthetics.

1.7 Necessity of High Speed Vision in Microrobotics

High speed vision is very demanding in traditional robotics. However, when facing a highly dynamic microworld, the necessity becomes much more substantial and even indispensable.

1.7.1 Automatic Control of A Microrobot

At microscale, the volumetric forces are negligible compared to the surface forces. A strong surface force exerted on an object of small mass will cause a huge acceleration. Thus, highly dynamic effects are ubiquitous because of the weak masses and inertia inherent to this scale. In this way, microrobotics researchers face different challenges from those of macroworld robotics. New sensors and processing have to be invented when dealing with such a highly dynamic world.



(a) (b)

Figure 1.16: (a) Visual controlled magnetic microrobot "MagPieR" performs a "2 mm dash" in the 2010 NIST IEEE Mobile Microrobotics Challenge, image courtesy of [Ivan 2011]. (b) A major bottleneck in framed based image processing: due to computational complexity, image has to be processed in a small region of interest thus restricts the user's workspace.

Two important issues in microrobotics should be addressed. The first is the *automatic control of a high speed microrobot*. The trajectory control of an autonomous

microrobot driven by a magnetic field in a constrained environment is presented in [Ivan 2011]. Fig.1.16(a) shows the microrobot "MagPieR" accomplishing a 2 mm dash in the 2010 NIST IEEE Mobile Microrobotics Challenge. The use of smart materials such as piezoelectric actuators provides adequate accuracy but introduces very high dynamics that cannot be easily controlled. Robust and ultra fast tracking of microrobots in large workspace is required in order to implement closed loop control schemes for such systems. The common working frequency of vision of 30 Hz is far from sufficient. The state-of-the-art visual tracking cannot ensure the proper monitoring of highly dynamic motion, thus confining micro-systems to simple tasks in a small workspace. Fig.1.16(b) shows a typical scene of a particle under high speed Brownian motion, with a limited resolution of 100×100 pixels. In all cases, highly dynamic phenomena due to low inertia must be recorded and tracked.

1.7.2 Teleoperated Micromanipulation

The second important issue is the *micromanipulation*. Numerous methods have been developed, based on rolling, push-pull or pick-and-place techniques [Xie 2009] [Haliyo 2003]. However, many problems still have to be overcome to build intuitive manipulation platforms. Only skilled operators can perform direct manual micromanipulation without any robotized instrument for the following reasons:

- The huge *scale difference* of position and force;
- The *fragility* of both the tools and the manipulated objects;
- The *complexity* of the various types of force fields;
- The high *sensitivity* to environment perturbations.

One solution is to build automated systems without human interference. Good results have been obtained by manipulating objects with a size of hundreds of micrometers with high throughputs and accurate positioning [Tamadazte 2010]. This manipulation mode fits the need of repeated tasks on a large number of objects. However, users do not receive enough feedback to learn from the manipulation.

Teleoperation is a common way for human to interact with the microworld. The user uses a *haptic* (literally means force sensation) device to control the position of the object and meanwhile feel the forces during the action. Commercialized force feedback haptic devices like Omega (Force Dimension), Phantom (Sensable) are widely available.

Haptic feedback overcomes the common difficulties in micromanipulation by providing an intuitive force feedback. The advantage of teleoperation over automatic microrobotic systems lies in that the user has interactions with the object being manipulated. He/she can use his/her expertise to take full control of the micromanipulation processes. It is said that automatic manipulation is suitable for repetitive tasks while teleoperation is better for individual tasks, in which the situation may vary from one time to another. Force feedback high resolution teleoperation has

assisted users in performing tasks that are either impossible or with high failure rate [Pacoret 2009][Xie 2011][Bolopion 2012]. Fig.1.17 shows a teleoperation experiment where a user obtains a force sensation of a contact between a gripper and a microsphere. Furthermore, Biologists and physicists can also profit from haptic devices to explore phenomena from their research fields on the fly.

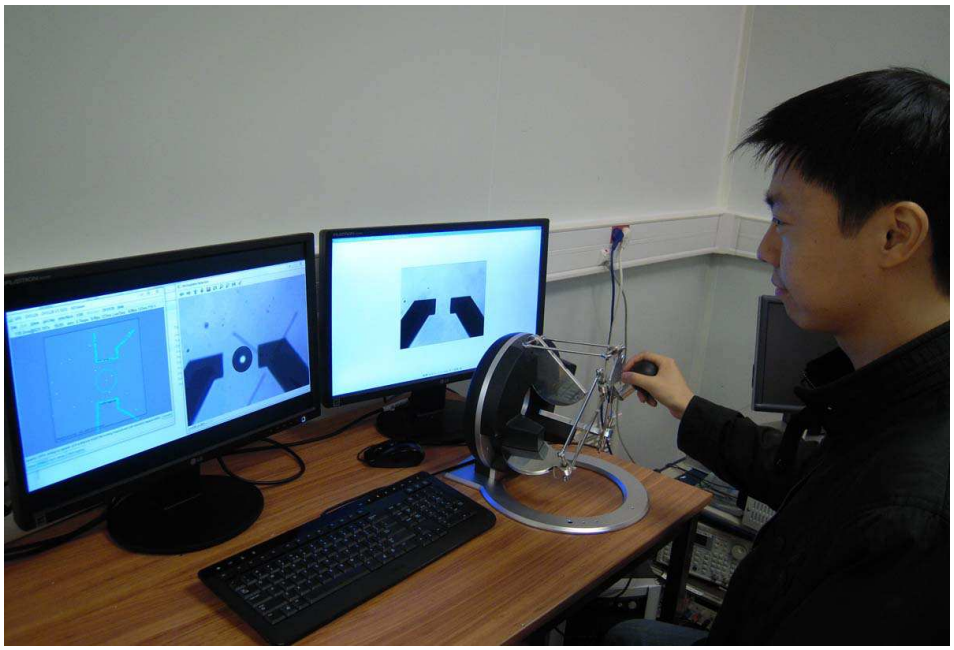


Figure 1.17: A user is tele-operating a microgripper with a haptic device (Omega, Force Dimension). Haptic feedback assists him to acquire an intuitive force sensation of the contact between the gripper and the microsphere.

It is easy to understand that the major problem in integrating haptic devices in micromanipulation systems is the lack of force feedback [Chaillet 2010]. Dedicated force sensors have been developed [Fahlbusch 1998][Beyeler 2009], but their installation inside micro-tools significantly increases the complexity and the cost of tool fabrication. In particular, even if some microgrippers offer sensing capabilities at the expense of a complex design [Duc 2008][Kim 2008], most still lack of force measurement capabilities [Lopez-Walle 2008][Andersen 2009]. Vision is a promising way to avoid the complexity of directly integrating force sensors on micro-tools. A high-speed microrobotic pick-and-place has been realized by visually recognizing the microgripper and the microspheres and by vision-based control [Zhang 2009][Zhang 2010]. To monitor the efforts during telemanipulation, visual information can be converted to force measurements by applying deformation models on the micro-tools after a calibration step [Cappelleri 2009] [Reddy 2010] [Chawda 2011], or be used to provide virtual assistance.

However, haptic force feedback faces a major challenge. The haptic system sampling frequency should be higher than 1 kHz in order to ensure the stability

of the haptic feedback and to achieve a real tactile perception [Colgate 1997]. The high dynamics in the microworld and the scale difference between the micro and the macro further strengthen this necessity. Due to these reasons, high speed vision stands out as a promising solution for haptic feedback micromanipulation.

Today, available vision sensors with desired specification, namely running at 1 kHz at least, are either record-and-then-replay or of low resolution. Moreover, image processing techniques able to match this frequency are very few and not robust to disturbances as discussed previously. Most of the existing vision-based microrobotic systems provide feedback at a couple dozens of Hertz, and their frequency depends on the size of the observed scenes [Greminger 2004]. The development of dedicated ultra fast vision sensors and associated algorithms will allow us to robustly track micro-objects in a large workspace at a sampling rate above 1 kHz and will lead to more pertinent and complex applications in this field of research.

1.7.3 Two Concrete Applications

This thesis addresses the issue of applying the event based vision in high speed haptic feedback micromanipulation by accomplishing two concrete applications, which mainly differ in the micro-tools employed. The first application is to develop a haptic feedback optical tweezers system that can be used to manipulate objects in liquid environments. The other application is to achieve stable haptic assistance by using a piezoelectric microgripper as the manipulation tool in ambient conditions.

1.7.3.1 Haptic Optical Tweezers

Optical tweezers are a revolutionary tool for micromanipulation and microscale force measurement [Grier 2003][Neuman 2004]. Micron-sized objects can be trapped under a tightly focused laser beam under a high numerical aperture microscope. The optical force is actually formed by laser's electromagnetic field, which produces stable trapping of dielectric objects. This technique is advantageously contactless and has been very popular in physics, cell engineering and other fields that need micromanipulation.

Since the laser may cause damages, silicon or polystyrene micro-beads are often used as a "handle" to indirectly manipulate the desired target, such as cells [Ashkin 1987]. The trapped bead can also be used as a force "probe" to touch the desired target to study biological or medical functions at microscale [Pacoret 2009][Haliyo 2003]. For example, to study how muscles produce work, Fig.1.18(a) shows two optically trapped microspheres attached to a filament of actin. A myosin molecule is then bound to the filament. The stroke of the binding is deductible from the position of the two trapped beads [Lewalle 2007].

If the optically trapped object is spherical, such as a microsphere, which is often the case, a simple spring model can be applied to describe the microscopic force exerted by the laser, such as:

$$F_{opt} = -K \times (P_{sphere} - P_{laser}) \tag{1.7}$$

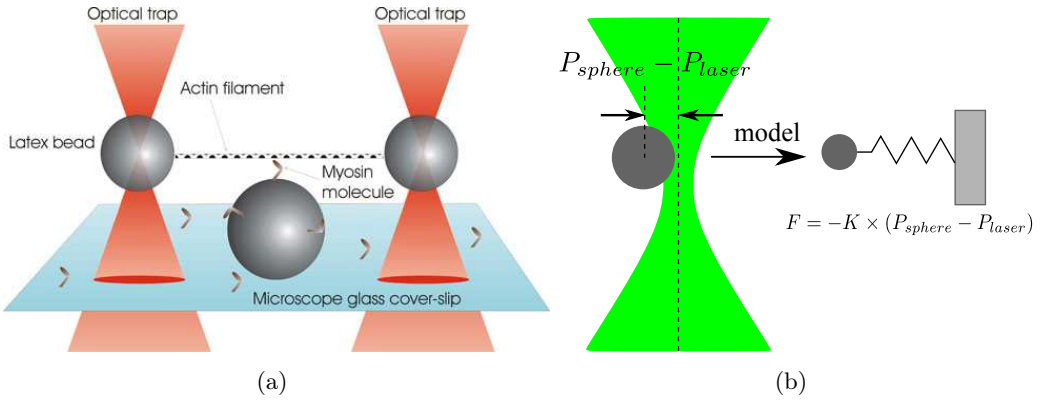


Figure 1.18: (a) A single filament of actin is stretched between two plastic beads held by optical tweezers. The myosin molecule can then bind to the filament, and its working stroke is measurable from the positions of the two trapped beads. Image courtesy of [Lewalle 2007]. (b) A spring model can be used to describe the position-force relationship of a trapped bead.

in which P_{sphere} and P_{laser} represent the sphere and the laser center position, respectively. F_{opt} is the optical force and K is the trap stiffness. The relationship is illustrated in Fig.1.18(b) as a spring intuitively. More profound mathematical and physical analysis on how the optical forces are generated by the laser gradient can be found in [Neuman 2004]. Under this model, the force measurement can be performed in an indirect way by precisely detecting the position of the trapped bead.

To couple the haptic system to optical tweezers, 1 kHz is required to ensure a realistic tactile sensation. A lower frequency will supplement a damp into the system, thus degrading the *transparency* between the macro and the microworld. In this case, the master part does not reflect precisely the interactions of the slave part, although the system may still be functional. Therefore, high speed position detection of microspheres at above 1 kHz becomes a fundamental issue in optical tweezers based micromanipulation.

Conventional image processing techniques cannot provide a plausible solution that tracks the positions of microspheres at this speed and in a complex micromanipulation environment. Therefore in Chapter 2, this problem will be resolved by employing a novel event based vision sensor to perform the needed tracking. The continuous Hough circle transform will be developed and the algorithm will then be validated by the detection of Brownian motion. Finally, the detected positions are transferred to a haptic device for the teleoperation of optical tweezers. Since our primary goal is to search a sound solution addressing the fundamental issue of high speed microsphere tracking, the experimental set-up is directly based on a formerly built system. The force feedback is constrained to two dimensional space in the first step.

Complex bio-medical applications require dexterous manipulations in three dimensional space. When displaced in the axial direction, the scenes frequently appear

defocused and blurred. Hence, the former continuous Hough transform is no longer adapted. Chapter 3 is driven by this new need of monocular 3D tracking. A robust circle fitting algorithm based on event data will be employed to estimate a bead's 3D location. The opto-mecanical experimental platform will be reconstructed to handle 3D actuation and force measurement. A multiple optical trapping facility will be added in the meantime to enable parallel teleoperation.

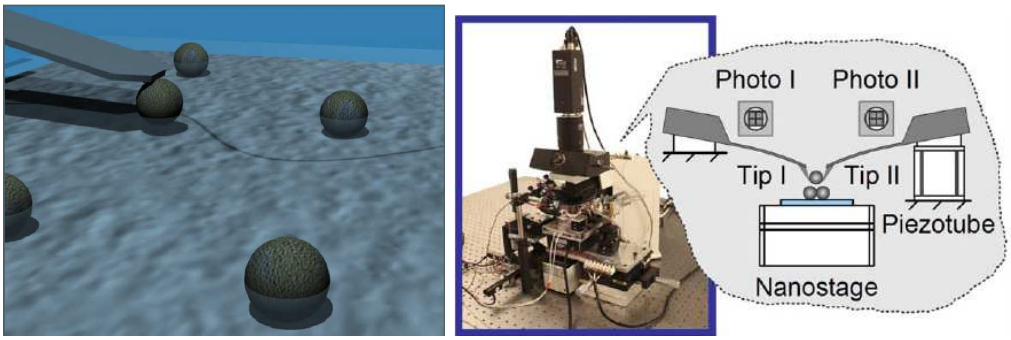
1.7.3.2 Haptic Virtual Assistance of A Microgripper System

The use of haptic feedback appears promising and benefits from the interactivity of manual manipulation, while facilitating it [Ferreira 2006]. In practical manipulation, the precise value of the force might not be interesting to the user and is not necessary for controlling micromanipulation procedures. Some position feedbacks obtained from vision sensors are sufficient as an enhancement or a guide to assist users in fulfilling the hand-on tasks, such as push, pick-and-place or even assembly. This set of problems is under the name of *haptic virtual assistance*. Instead of transmitting the exact micro- or nanoscale forces, haptic feedback serves just as an operation assistance. These applications do not require the existence of a precise force-position model on the micro-tool, thus the haptic feedback can be coupled with almost every types of end-effectors that operate in either ambient or liquid conditions, e.g. magnetic tweezers, AFM and microgripper.

As an application example in [Ammi 2007], haptic feedback is used to keep the user's motion on a specified path by creating virtual "walls" that actually do not exist, see 1.19(a). Any attempt to surpass the desired trajectory will be hindered or blocked by the "walls" to avoid any dangerous operation. The same idea also frequently appears in the so-called comanipulated surgical robots. Unfortunately, these examples are mostly limited to proof of concepts. For the second application of the thesis, the set of problems about tracking the positions of a micro-tool to provide haptic virtual assistance for micromanipulation will be tackled.

A major problem for this type of application is to guarantee *stability*. A theoretical analysis of the stability in haptic coupling schemes for micromanipulation has been presented in [Bolopion 2009]. It is a well known fact that tools and objects at microscale have negligible volume forces and are not subject to the phenomena of inertia. The commencement and the termination of a movement are instantaneous. In contrast, the haptic interface and the human arm do have inertia. The differences in the mechanical properties between the master and the slave systems are the sources of instability. Oscillations induced by altering energy transferred from one part of the system to the other can easily appear in a system with an insufficient sampling rate. If these oscillations are important or even divergent, they will finally impede the operator from performing the correct gestures rather than assisting him. Optical tweezers do not suffer as much from the stability issue, because it is intrinsically a first order damping system.

Stable virtual guides for pick-and-place operation of microspheres based on two atomic force microscopy cantilevers has been developed [Xie 2009][Bolopion 2012].



(a) (b)

Figure 1.19: (a) Haptic feedback serves as a virtual guide to keep the user's motion on a specified path. Image courtesy of [Ammi 2007]. (b) Virtual guides for pick-and-place operation of microspheres based on two atomic force microscopy cantilevers. The laser beams have to be aligned with the cantilevers and photodiodes. Image courtesy of [Bolopion 2012].

It uses an AFM composed of two independent cantilevers, which made the system very difficult to use. The cantilevers were equipped with force sensors (two optical levers) that provided haptic feedback, but this increased the complexity of the setup since the laser beams had to be precisely aligned with the cantilevers and photodiodes, see Fig.1.19(b). It was thus definitely not a system that could be used by non-expert users. For these reasons, a simple piezoelectric gripper is chosen. Compared to [Bolopion 2012], where the two AFM cantilevers were equipped with photodiode sensors, here the event-based vision sensor and algorithm are employed to provide the required feedback, which is embedded sensor deprived and much easier to use.

In Chapter **??**, it will be shown that the integration of the event based vision sensor is a suitable solution for monitoring the high dynamics common at microscale and ensuring the stability of the haptic feedback. This time, however, the shape of the micromanipulator is more complex and arbitrary, not limited to circular forms. An original and dedicated algorithm – the event based shape registration – will be designed to perform a visual shape tracking of the piezoelectric microgripper for haptic virtual assistance, and that, much faster than the required 1 kHz to guarantee the stability of the system.

Chapter 2

Asynchronous Event Based 2D Microsphere Tracking

Contents

Silicon retinas are especially suitable for tasks like tracking, because only dynamics are recorded and the generated data reflects only moving edges. To begin with the simplest example, this chapter mainly focuses on the detection and localization of objects of circular forms with a known radius, targeting for the optical tweezers application. Firstly, the event based continuous Hough transform algorithm will be introduced. Applications of high dynamic Brownian motion detection and 2D force feedback teleopereated micromanipulation of optical tweezers will then be demonstrated. The applicability of the algorithm can be easily extended to other robotic visual tracking tasks.

2.1 Reliable Haptic Optical Tweezers

Optical tweezers now enter new era for micro/nanomanipulation in the biological or medical field. The Gaussian beam profile of optical tweezers generates gradient

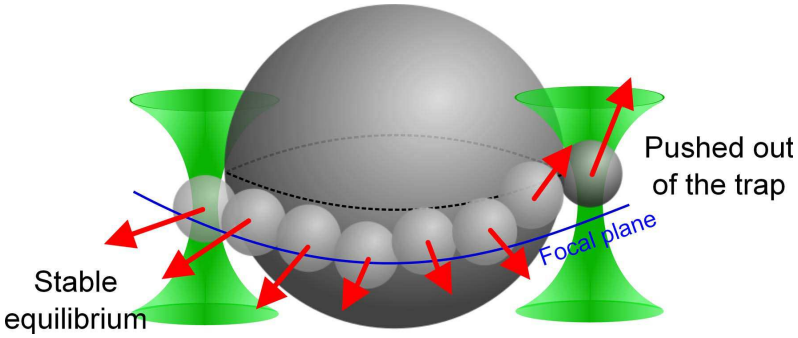


Figure 2.1: A small microsphere (the probe) is moving along the surface of a big microsphere. Initially in an equilibrium position, the probe moves from below the medium height of the surface upward. The combinational force finally drives the probe out of the optical trap due to weak trap stiffness in z axis.

forces that are strong enough to trap a microsphere. The sphere will then serve as a "handle" to manipulate other samples, e.g. DNA stretching. Remember that in Section 1.7.3.1, we show that conventional force estimation routine concerns to apply a simple spring model on the optically trapped sphere. In this way, locations of particles with respect to the laser center become valuable information for in situ force measurement.

Equipped with this type of force sensing technique, optical tweezers becomes an even more powerful tool for micromanipulation. The trapped microsphere can now serve as a force "probe" to examine the physical characteristics of a target object, such as the stiffness of cells, thus enabling a broad variety of potential applications, especially in biology or medical science. In this thesis, "probe" always refers to the microsphere trapped by optical tweezers. Fig.2.1 shows a concept application of surface probing. Forces can be felt by the user when he/she conducts the probe to skim over the surface structure of the target object. The force direction felt by the user is perpendicular to the tangent plane of the target surface, thus allowing one to acquire very intuitive interactions with the form and the stiffness of the exploring micro-object. As can be seen, the key issue to open this type of applications lies in the precise position detection of the trapped probe.

The position detection faces two major challenges. Firstly, force feedback through a haptic interface requires a high sampling frequency of 1 kHz to achieve a realistic tactile sensation. Lower sampling frequency will artificially increase damping into the system thus degrading the transparency of the haptic feedback. Secondly, complex bio-medical manipulation environment further demands robust measurement methods. Immunity to environmental noise is an important factor in order for the interactions be reliable regardless of external perturbations (obstacles, defocusing, impurities, shadows...). Therefore, high speed and meanwhile robust position detection of microparticles becomes a fundamental issue in optical tweezers based micromanipulation and is the major contribution of this chapter.

In this chapter, we concentrate on resolving the issue of the high speed tracking

of microspheres. Several experiments, such as Brownian motion detection, will be conducted to validate the algorithm and in the meantime demonstrate the tracking capability. Finally, the force feedback will be applied on optical tweezers. The force is temporarily limited to two dimensional space since an existing 2D experimental setup is reused to quickly show the plausibility and the potentiality of the employed method. A more powerful and complete optical tweezers system will be the subject of the next chapter.

2.2 State-of-the-art of High Speed Microparticle Tracking

The state-of-the-art will be introduced from two aspects. Different tracking devices will be firstly compared and then the tracking algorithms.

2.2.1 Position Detection Devices

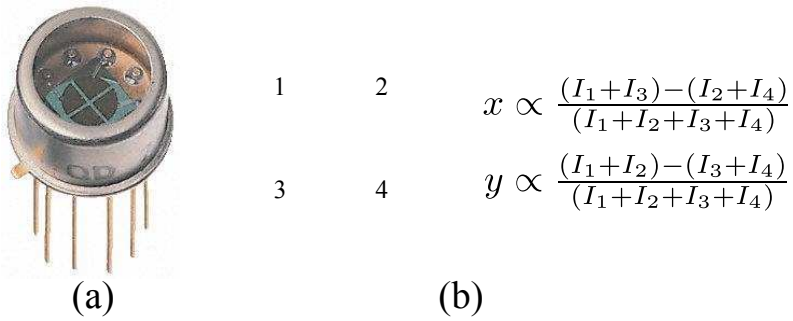


Figure 2.2: (a) A photo of quadrant photodiodes. Image courtesy of Radiospares. (b) The working principle of quadrant photodiodes. The difference of output current between the left and the right halves gives the position of the bead on the x axis, while the difference between up and down halves gives the position on the y axis.

In optical tweezers sensing techniques, few vision sensors allow high speed position detection. Quadrant photodiodes (QPDs) achieve more than 1 MHz detection bandwidth (Fig.2.2.1(a)). Initially, a calibration step is required to center the bead on the photodiode plane so that the output current generated from each quadrant cancels each other. During measurement, the slight movement of the bead within the range of QPDs will change the current in each quadrant. The difference of current between the left and the right halves gives the position of the bead on the x axis, while the difference between up and down halves gives the position on the y axis, see Fig.2.2.1(b). This method thus requires the scene to be highly clean. Any asymmetry condition, e.g. obstacles and impurities, during the measurement process will influence the output current thus is strictly not tolerated. Furthermore, QPDs limit the working distance and require strict sensor alignment.

Multiple object tracking with complex robust processing is only possible with video cameras. Microparticle position measuring with video cameras now gradually replaces QPD in the most recent publications. However, modern high speed cameras suffer the same constraints in microparticle tracking as in general tracking tasks that have been explained in detail in chapter 1. As the state-of-the-art of microparticle tracking techniques, ultra high speed cameras remain an offline solution due to large amounts of data stored in a local memory during acquisition [S.Keen 2007]. Rapid CMOS cameras have been developed allowing a reduction of the region of interest. They can transmit the image in real-time, but the entire frame online processing remains time consuming [Gibson 2008]. Dedicated image processing hardware (e.g. FPGA) [Saunter 2005] or "smart" CMOS cameras with on-chip processing [Constandinou 2006] [Fish 2007] or GPU accelerated technology [Lansdorp 2013] are used, but these cost extra expenses, elongate development cycles and require specific hardware knowledge. Current particles tracking software is mainly based on centroid.

Table 2.1 gives a summary of the prices and experimental performances of the vision systems described above for micro particle tracking; it also shows the clear advantage of the use of asynchronous event based silicon retinas.

Type	Price (K€)	Acquisition (KHz)	Processing (KHz)	Comment
Photodiode	0.2	10*	10*	Only single sphere, difficult to align
Ultra-fast cam	30	35+	offline	Not for real-time use
CMOS	1.5	2	0.5	Image processing with a PC
CMOS +FPGA	3-4	2	2	Special hardware knowledge required
DVS	2	1000*	30**	High speed and low consumption

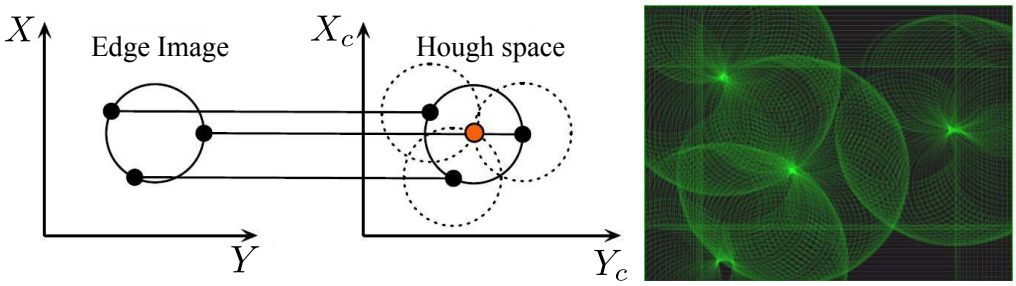
Table 2.1: Comparison of fast vision solutions in micromanipulation. *These frequencies are the information refreshing rate, rather than classical cameras' frame rates. **to be achieved in the thesis.

2.2.2 Candidate Algorithms

To find circles in images from a video camera, we mean to extract the parameter pair $(x_c, y_c)^T$ or triplet (x_c, y_c, R) describing the center and the radius of circles. Typical ways of searching circle parameters will be studied in this section, which are Centroid, Histogram, Template Correlation and Hough Transform, etc.

The centroid method is the simplest method which has already been explained in Chapter 1, Equation 1.1. In histogram methods, one calculates the sum of the pixels' gray levels in the y axis for each location of x. The same applies for each coordinate of the y axis. So the circle center will appear to be the peak location in histograms of

both the x and the y axes. The idea of cross-correlation is to use a template image to match a part of the image being processed. The maximum of the correlation indicates the interested location where the template and the image in question have the most similarity. In [Gosse 2002], to measure the bead displacement since the last frame, a small image centred on the previous particle position is used. To maximize a cross-correlation function gives the bead position, which allows sub-pixel resolution.



(a) (b)

Figure 2.3: (a) A brief illustration of Hough transform. Each point on the edge of a circle in the edge image votes a circle in an accumulation space, also called Hough space. The peak of Hough space signifies the circle center location. (b) A Hough space generated by using synthetic images. Several peaks correspond to several circles centers from the edge image.

It appears frequently in real-world scenes that the target circles overlap or collide with other objects. To be able to tackle noise and outliers, Hough transform is such a well validated and the most used circle detection algorithm in image processing [Hough 1959].

The conventional Hough transform consists in two main steps. For easier illustration of the concept, let us suppose that the radius is known to users and only the circle center $(x_c, y_c)^T$ is under quest. The input to the algorithm should be edge images. Hough transform requires to transform the unknown parameters in the original image into a local maximum in another space. To illustrate, each point on the edge of a circle votes an entire circle in an accumulation space, also called Hough space. Each full circle in Hough space thus corresponds to each valid pixel in the edge image, see Fig.2.3(a). Once the accumulator space is constructed, the second step includes finding peaks by looking for local maxima. Fig.2.3(b) shows the hough space generated from synthetic image data. The strong peaks correspond to multiple circle centers and are isolated by setting a threshold. The thresholding results in small blobs of pixels in the peak area. The x and y values of the pixel locations are then averaged to approximate the circle center to achieve sub-pixel resolution.

If the circle radius is unknown then a supplementary parameter should be introduced. The Hough space can theoretically be extended to 3D. In practice however, this direct modification is too time consuming and is avoided by separating the

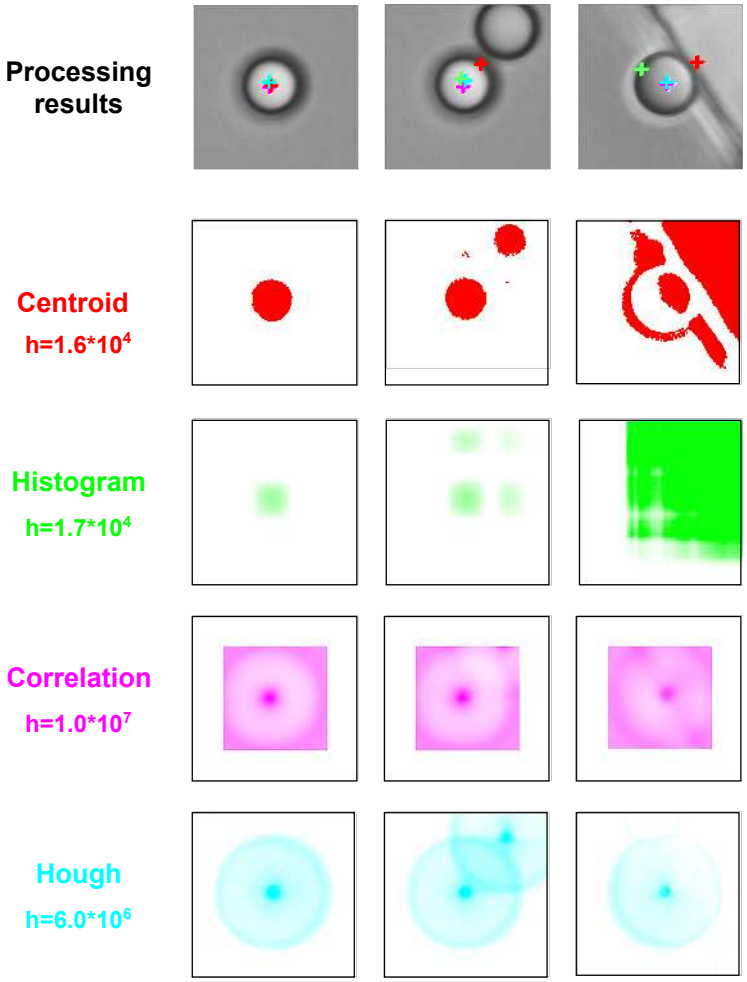


Figure 2.4: A comparison of typical circle detection algorithms. Centroid and Histogram are fast, but are also very sensitive to noise and outliers. On the other side, Correlation and Hough transform are more computational expensive but more robust to unconditioned environment. The value h indicates the calculation steps for processing a 128×128 gray level image with circle radius of 25 pixels. We see that the last two methods are two orders more complex than the first two.

procedure into two successive lower dimensional procedures. [Ioannou 1999] has presented a method by first applying 2D Hough transform for center localization and then a 1D radius histogramming for radius validation. This chapter focuses only on two dimensional circle tracking because the force feedback is constrained to the xy plane in the first step.

Centroid and Histogram are fast, but very sensitive to noise and outliers. Fig.2.4 compares the speed (in number of calculation steps for processing a 128×128 gray level image with circle radius of 25 pixels) and robustness of the above-mentioned algorithms. The first row are tested images superposed by the detection results. Centroid and Histogram methods are two orders faster than Correlation or Hough,

but they are not able to track multiple adjacent objects (Fig.2.4, column two) or adapted to contaminated scenes where obstacles lie (Fig.2.4, column three). On the contrary, the other two methods are able to handle the difficulties by significantly sacrificing the processing speed. In circle detection, Hough is usually more robust than correlation because only circular forms are taken into account and tracked. As a result, we choose to use Hough transform as a starting point for microsphere position detection.

2.3 Microsphere Tracking Using DVS

The traditional circle detection process is computationally expensive since it first involves an edge detection procedure and then Hough transform. We propose a novel event based Continuous Hough transform that uses DVS's events as input to track microspheres, a type of tool immensely used in optical tweezers systems. DVS's sparse input opens a different point of view for algorithm design and enables high speed and meanwhile robust tracking method.

Two groups of experiments are conducted in order to validate the algorithm, to demonstrate the tracking capability and to analyse the performance. They are: (1) to track multiple microspheres simultaneously transferred by a microstage; (2) to detect Brownian motion of microsphere as a live demonstration of high dynamic motion detection in the microworld. Brownian motion can eventually feedback to users or be further processed to calibrate the stiffness of the optical tweezers.

2.3.1 Event Based Continuous Hough Transform

The sparse data produced by DVS and their temporal asynchrony allow to run the dedicated algorithm in real-time at extreme high speed. The mathematical derivation of Continuous Hough Transform is as follows.

Let us start from the case where the radius R is fixed or known to users, so only the center coordinates $(x_c, y_c)^T$ must be determined. Remember that S_t represents the set of events active at time t and are stored in a First In First Out (FIFO) data structure. The event-based Hough transform using asynchronous events can then be defined as:

$$H : S_{t_k} \rightarrow A_k = H(S_{t_k}) \tag{2.1}$$

where A_k is, in the case of two parameters search, an integer matrix satisfying :

$$A_k(i,j) = \begin{cases} = 1 & if\ (x-i)^2 + (y-j)^2 = R^2 \\ = 0 & otherwise \end{cases} \quad for\ \forall i,j \in [0,128],\ (x,y) \in S_{t_k} \tag{2.2}$$

The Hough space voted by all the incoming events in the FIFO has been expressed as $H(S_{t_{n-l+1},t_{n+1}})$, in which t_{n+1} is the current time and l is the FIFO size. The general algorithm is given using the following recursive formula, under which

lies the main advantage of Continuous Hough over conventional ones:

$$H(S_{t_{n-l+1},t_{n+1}}) = H(S_{t_{n-l},t_n}) + H(S_{t_{n+1}}) - H(S_{t_{n-l}}) \tag{2.3}$$

with the convention that $H(S_{t_{n-l},t_n})$ is the sum of the vote in Hough space by events generated during (t_{n-l}, t_n). The update of location by event based Hough is continuous. When the newest event is coming into effect, the last one in the FIFO is dropped correspondingly.

Afterwards, the circle center $(x_c, y_c)^T$ can be detected simply by selecting the local maximum in Hough space. In applications where the object movement is small, like the Brownian motion detection, the resolution is generally too low for a precise analysis. In order to achieve sub-pixel resolution, a center-of-mass algorithm (or Centroid) within the neighbourhood of the peak in Hough Space has been calculated instead of choosing only the local maximum.

More practical issues may be considered in real applications. A region of interest (ROI) can be set for one particle if the scene presents multiple circles. Moreover, to dynamically track all the moving particles, their appearance or disappearance should also be detected by setting validation thresholds and ensuring a boundary checking during each iteration. The event-based Continuous Hough circle transform is summarized in Algorithm 1.

Algorithm 1 Event-based Hough Circle Tracking with fixed radius

```
Initialize the parameters
while an incoming new event e(p, t_{n+1}) do
    Apply equation (2.3)
    if (x, y) ∈ one of the ROIs then
        Update the center of the particle.
    else if (x, y) ∉ ROIs then
        Detect a new local maxima close to (x, y) for possible new particle.
    Centroid in Hough space for sub-pixel resolution.
    Threshold validation and boundary checking.
```

2.3.2 Multiple Microsphere Tracking

In this section, preliminary experiments of detecting multiple microspheres displaced by a piezoelectric microstage have been conducted to validate the tracking algorithm. Using DVS's input, the asynchronous continuous Hough Transform can achieve an information refreshing rate upto 30 kHz.

2.3.2.1 Setup

The experimental setup is plotted in Fig.2.5. The light beam is projected from above and the microscope objective (×20, NA.0.25 or ×100, NA.0.8, Olympus) is placed upside down to observe samples in the Petri dish. The Petri dish is placed inside

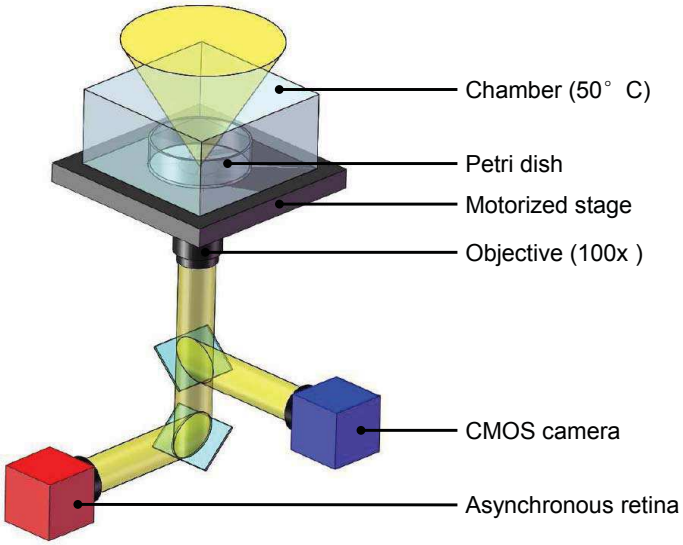


Figure 2.5: A brief illustration of our experimental setup.

a chamber equipped with thermal-coupled temperature controller. The platform is a controllable micro-stage driven by a DC-Servo Motor Controller (C-843, Physics Instrument). The light beam is split into two rays and redirected separately to DVS and to a classic CMOS camera (Basler). The two cameras are aligned to produce the same view of the microscope. The CMOS camera is used to provide a visual feedback to the user for system control. Polystyrene microspheres (Polysciences) of sizes 6µm, 10µm and 20µm are placed in pure water for tracking.

2.3.2.2 Experiments

Experiments are separated into two parts. The first part demonstrates the capability of simultaneously tracking multiple microspheres by Continuous Hough while the second part analyses the real-time processing speed.

In the first part of the experiment, thousands of microspheres with a constant diameter of 10µm are tracked under $\times 20$ magnification. The field of view contains a mean value of 50 microspheres moving simultaneously. DVS sensibility is adjusted manually to our experimental condition for a better detection.

In Fig.2.6(a), a screen view shows a frame obtained by summing events over a time window. The FIFO size is set to a maximal size of 2000 events. As shown in Fig.2.6(a), the algorithm is capable of detecting several dozens of moving microspheres simultaneously. It is important to point out that only the activities of pixels are taken into account, therefore the polarity does not have an impact on the computational process. The ground truth envelope corresponding to the number of microspheres manually counted every 5 seconds (Fig. 2.7) shows that the method does not suffer any loss. A frame-based Hough using accumulation maps of events from the silicon retina has been implemented to provide a comparison.

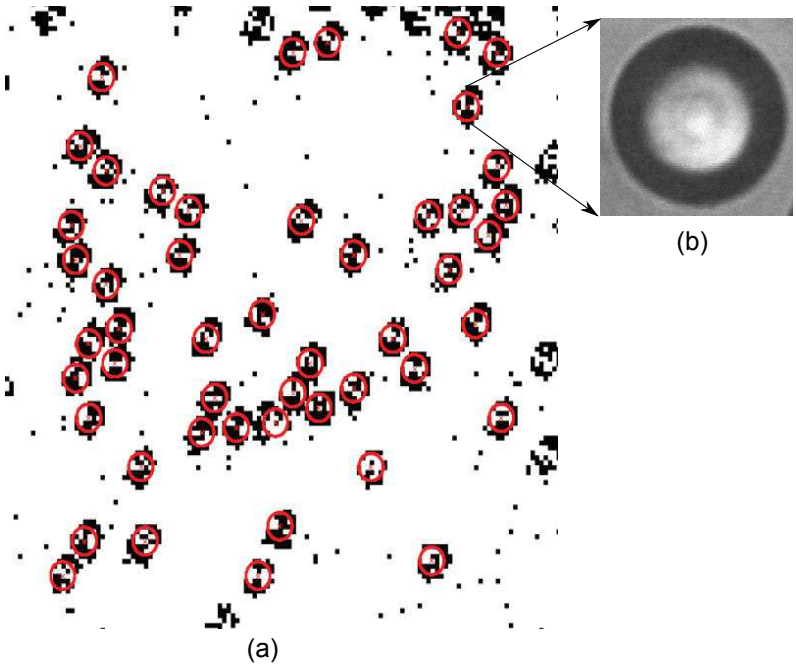


Figure 2.6: (a) A multiple microspheres (10µm) detection example. Black dots are the events accumulated in the FIFO. The red circles are the detected microspheres. (b) An image of one microsphere taken by a conventional camera.

The sudden drop of circle number by frame-based Hough is due to the irregular motions of the microspheres. When the micro-stage stops or accelerates, it causes a sudden burst of noise in images. As a result, the frame-based Hough technique cannot handle without a robust tracking filter. The event-based Continuous Hough overcomes this problem thanks to its "update-from-last-position" mechanism, which acts intrinsically as a tracking filter.

Further experiments are fulfilled to examine the time consumption of the algorithm. The program implemented in C++ under Windows is running in real-time. The microspheres used were of a known radius of 6µm while the magnification was set to ×100. The results of real-time experiments are shown in Fig.2.8 where three microspheres are moving in the water current. The position refreshing rate is shown in Fig.2.9. It shows that the average time for processing an event is around 32.5µs, namely a 30kHz frequency. If the events are distributed over three circle edges, the detection of each microsphere's position can then be computed at a mean frequency close to 10kHz.

It is important to emphasize that no specific hardware was used except for DVS. The experiment relied on an off-the-shelf 2.9GHz Dual core desktop PC for the whole system including the microstage control, with a total CPU load little more than 50% of its power and a memory consumption of about 4MB. If compared to frame-based techniques, in terms of rapidity, memory consumption and hardware price, event-based vision outperforms frame-based vision by far. It is also important

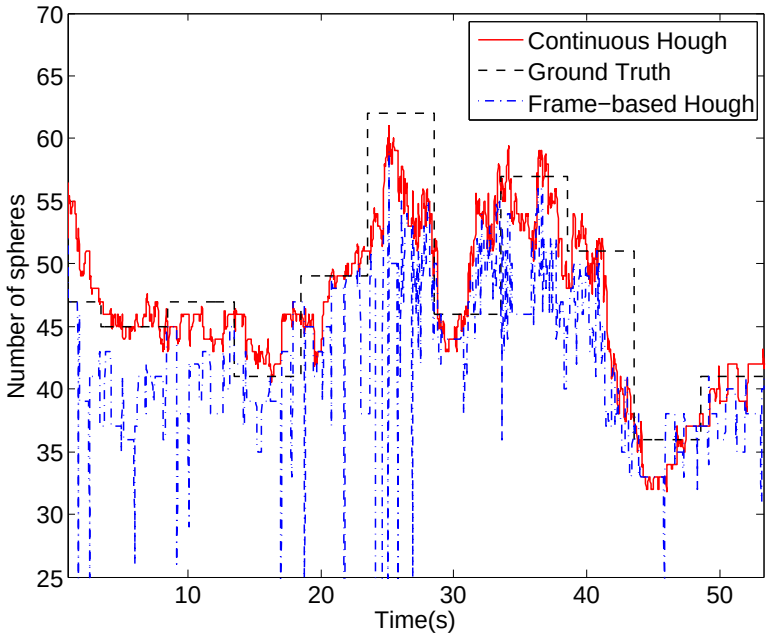


Figure 2.7: This figure compares the number of microspheres detected over time using the frame-based (blue) and the event-based Hough (black) transform with DVS. The microspheres are counted manually every 5 seconds (the red envelope).

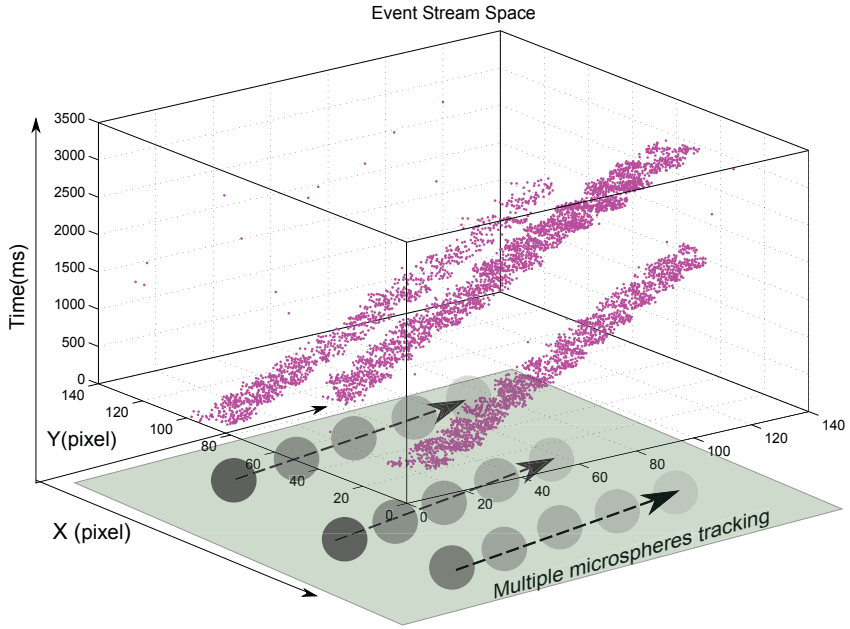


Figure 2.8: Event space of fast moving circles driven by water current. The vertical axis represents the time, the other two are the xy plane. Noise can be observed, however our algorithm is robust enough to overcome temporally unrelated noise.

to point out that the used event-based method is a much more complex algorithm than the classical center-of-mass method. Compared to pure tracking of events that have been reported in [Delbruck 2007], the computational load of the Hough method is higher (5 times) as it is aimed to detect specific shapes while in Delbruck et al, the algorithm is tracking blobs of events regardless of shapes. Introducing a constraint on the tracked shapes ensures that in the case of several moving objects in collision, only spheres will be correctly located. This introduces robustness in the measurements as is usually needed in real-world micromanipulation, notably, in the application of teleoperated optical tweezers.

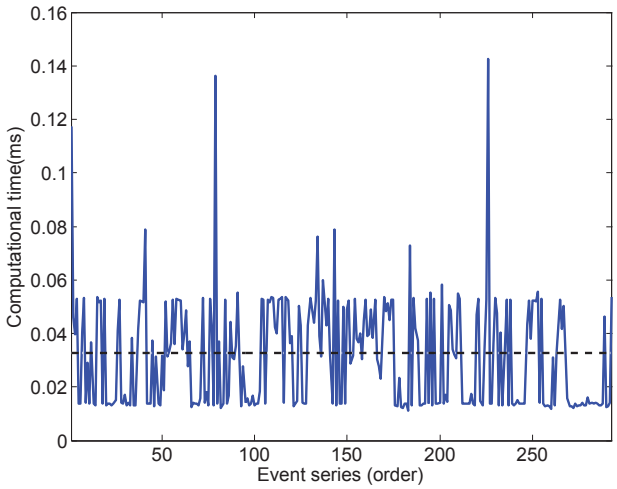


Figure 2.9: Time consumption in real-time test. The mean time for processing an event is at about 32.5µs (dot line), namely a 30kHz refreshing rate. The peaks are caused by the OS interruptions.

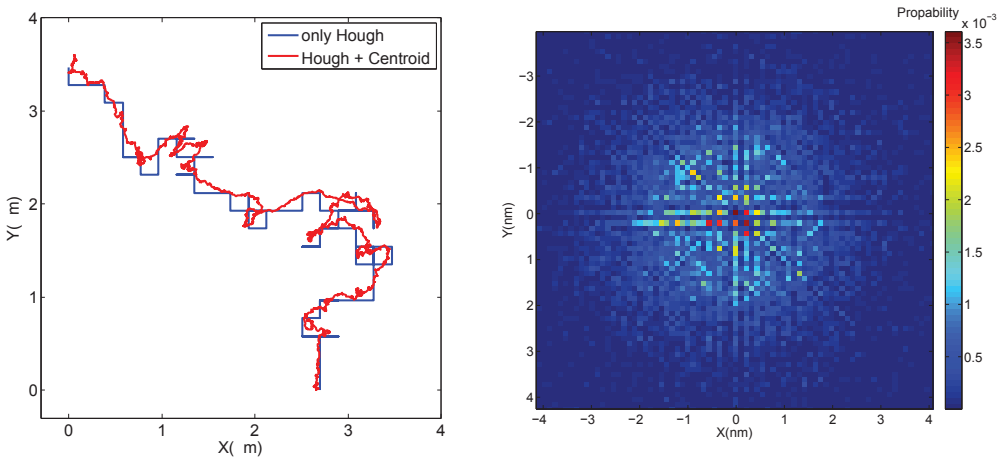
2.3.3 Brownian Motion Detection

Brownian motion corresponds to the thermal agitation of micro/nano sized particles in the fluid. The motion appears to be small and fast in the scene. This is definitively an example of highly dynamic phenomena in the micro-world, which provides a good chance to show the up limit of our high speed tracking capability. More importantly, Brownian motion can be fed back to operators in teleoperated systems that allows to sense the dynamics of the microworld and the position data of Brownian motion can be used to calibrate the stiffness of the optical tweezers.

This experiment concerns the detection of Brownian motion of microspheres within size of 2.7µm. The temperature is controlled at around 50 °C in a thermal controlled chamber.

The theoretical displacement of the particle satisfies a centered normal distribution. The standard deviation of the distribution depends on temperature, viscosity and the sampling frequency in a synchronous sampling mode [Crocker 1996]. In

our condition, 70% of the measurements are smaller than 2 nm. The position signal behaves like a white Gaussian noise and has a large frequency spectrum. This phenomenon is ideal to test the sensitivity and dynamics of the presented system.

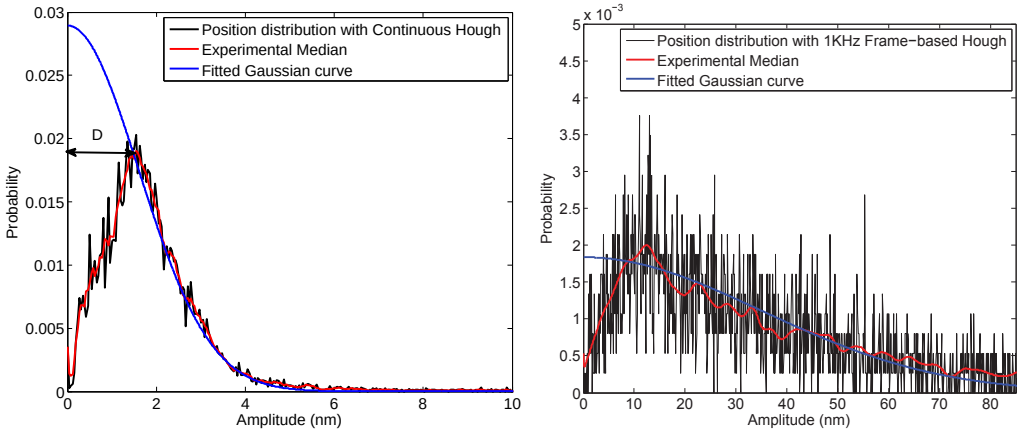


(a) (b)

Figure 2.10: (a) Comparison of the detection results (Cartesian coordinates) by using only Continuous Hough and by combination of Hough and Centroid. With only Hough, the result has pixel accuracy, while with the later method, a sub-pixel resolution can be achieved. (b) The displacement $(\Delta x, \Delta y)^T$ distribution of a microsphere's Brownian motion. This is a top view of the Gaussian distribution with a standard deviation $\sigma_x = 1.53\ nm, \sigma_y = 1.59\ nm$. The color axis represents the probability counts in each location. Note that the result is more sensible around the voting angle in Continuous Hough Transform.

A 2.7µm microsphere's trajectory is plotted in Fig.2.10(a) with and without the Centroid algorithm applied. The precision is improved by the use of the Centroid algorithm, it allows a sub-pixel accuracy (1 pixel = 0.19 µm, for the ×100 objective). Fig.2.10(b) shows the spatial distribution of the microsphere's positions detected by the Continuous Hough. The pattern results from an angular sampling step of 5 degrees used for the Hough accumulator. The chosen sampling value is the best experimental compromise for rapidity and precision. Fig. 2.11(a) shows the amplitude $\Delta r = (\Delta x^2 + \Delta y^2)^{1/2}$ of the Brownian motion measurements compared to the theoretical distribution. The standard deviation of our measurement along x, y axes is $\sigma_x = 1.53\,nm, \sigma_y = 1.59\,nm$. An incomplete Gaussian is obtained since too small displacements cannot be optically measured. The lower boundary to consider the measurements as being accurate is equal to the maximum of the distribution shown in Fig.2.11(a) (which is experimentally measured to be $1.9\,nm$). Fig.2.11(b) shows the distribution calculated by the frame based Hough for comparison. Frames are obtained by accumulating events to simulate a high speed CMOS camera of frame frequency 1 kHz. By applying a Centroid after Hough, the method is able to track a single microsphere with sub-pixel resolution offline. The measured standard devi-

ation of the frame based Hough in Fig.2.11(b) is 35 nm. Compared to Fig.2.11(a), this deviation is significantly larger due to the lower sampling frequency. The distribution also has larger variance since less temporal correlation information is used by the frame-based Hough.



(a) (b)

Figure 2.11: (a) The relative particle displacement distribution $(\Delta x^2 + \Delta y^2)^{1/2}$ (in black) and the fitted Gaussian envelope (in blue). The red curve is the experimental median. The camera does not reliably detect low displacements below 0.01 pixel (D). The standard deviation of the fitted Gaussian curve is 1.6 nm. (b) The same displacement distribution obtained from Frame-based Hough. Frames are obtained by accumulating events to simulate a high speed CMOS camera of frequency 1 kHz. The centroid algorithm is applied in successive of Hough. The standard deviation of the fitted Gaussian curve (blue) is 35 nm.

For real-time tracking, the bandwidth of DVS is not limited by any frame rate constraints due to its asynchronous principle. The temporal resolution limit is set by the minimum timestamp which is 1 µs and the pixel bandwidth which is more than 3 kHz. The DVS Brownian motions' measurements have a mean temporal resolution of 2.15 µs. Concerning the system latency, events are sent in the form of packets through the USB interface. By choosing an optimal device interrupt polling interval, the latency of the system can be achieved at the level of 2.8±0.5ms [Delbruck 2007]. This is fast enough for the control of microsystems in closed loop [Ivan 2011].

The selection of events in the circle's region of interest allows eliminating the influence of the fixed-pattern noise. The Hough algorithm itself is also very robust to noise, because only the dominating value in Hough space will be taken into account for the final decision of the sphere center. As a minor default, the centroid calculation used to achieve sub-pixel resolution will indeed bring back slightly some noise if the height of the peak region is not discriminating. Fortunately, this problem can be fully resolved by the circle fitting method presented in Chapter 3.

In this section, first tracking examples are conducted to demonstrate real-time multiple circle tracking with DVS at a refreshing rate up to 30 kHz and its application to the detection of Brownian motion. The asynchronous event-based acquisition and derived methods outclass existing micromanipulation methods relying on CMOS or ultra-fast cameras. Although the presented method has been developed in pure software running on a standard PC platform, its implementation in embedded hardware will reach even much higher performances. They have substantial space for innovation and the potential of realizing small, fast, low power embedded sensory-motor processing systems that are beyond the reach of current frame-based techniques.

2.4 2D Haptic Feedback Micromanipulation with Optical Tweezers

The great effort we have made to track microspheres is due to its essential role in optical tweezers systems. A microsphere is trapped and then serves as a "handle" or "probe" to perform manipulation of other objects and to measure forces. Teleoperation of optical tweezers is achieved by coupling with a haptic interface. This provides users direct and intuitive microscopic interactions. The major difficulty to generate a real and reliable force feedback lies in its high frequency requirement of more than 1 kHz.

We have achieved the goal by using an event based sensor and the presented circle tracking algorithm. In this section, the result of DVS will be compared to that of a conventional frame based one to show the advantages of our technique. Complex tasks of object touching and surface exploration is performed demonstrating the robustness and efficiency of the algorithm. This is the first time microspheres are used to touch targets of arbitrary form and color. The experimental setup is based on an existing 2D platform, so the force is currently limited to two dimensions.

2.4.1 Strategy of Haptic Coupling with Optical Tweezers

In a teleoperated system, the haptic interface is usually referred to as the master device and the micromanipulator as the slave device. The first bilateral coupling of a teleoperation system to optical tweezers is reported in [Arai 2000]. The meaning of "bilateral" in the coupling of the micro- and macro-world can be explained as follows:

- Position coupling streamline: The position orders P_{hand} are received by the haptic interface, scaled by G_{pos} and then used to command the microstage displacement to generate laser movement P_{laser}.
- Force coupling streamline: At mean time, the microscopic forces F_{opt} detected by the vision algorithm $H(S_t)$ are scaled by G_{force}. The magnified force F'_{opt} is then fed back to the user through the motor mounted on the haptic device.

Our strategy of haptic coupling with optical tweezers is to use homothetic gain, namely G_{pos} and G_{force} are constants. The coupling scheme is illustrated in the block diagram of Fig.2.4.1.

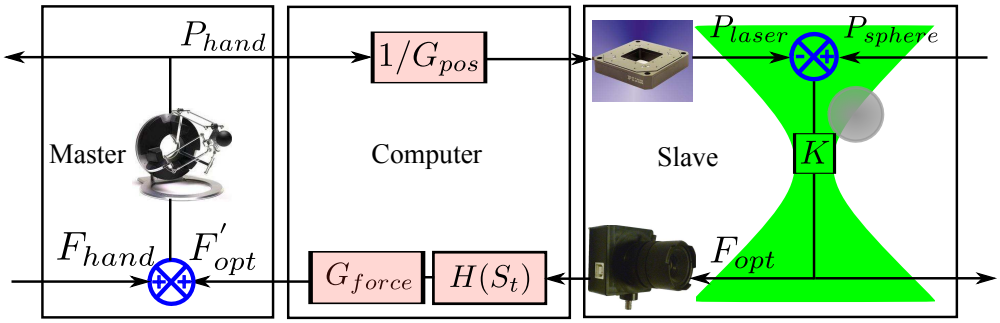


Figure 2.12: A schematic diagram of the bilateral coupling of position and force in a teleoperated optical tweezers system.

In [Arai 2000], the force feedback frequency is too low to provide operators with a reliable sensation. Double-trap [Onda 2012b] or multi-trap [Onda 2012a] parallel systems are developed, however the bandwidth problem of such haptic feedback remains unresolved. We face the challenge by proposing for the first time a 1kHz event based vision solution dedicated to force feedback on optical tweezers.

The estimation of contact forces in conventional frame-based techniques during touching procedure requires complex image segmentation algorithms that can only operate at low processing speeds. In [Pacoret 2009], the haptic coupling is conducted in a similar way as illustrated above. An exploration task has been performed successfully in a tolerable speed for guiding a sphere along a dark flat wall or a corner. This touch experiment requires only a simple image thresholding technique to eliminate objects of dark color. However, the method employed in this chapter aims for the touching or exploring of objects of various shapes and colors, where simple image segmentation algorithms do not work.

2.4.2 Haptic Feedback Optical Tweezers System Setup

The 2D system is composed of conventional optical tweezers that has been constructed by a previous Ph.D student: laser source (532nm, 300mW) and oil-immersion high numerical aperture objective (UPlanFLN 40x, NA 1.3, Olympus). A single laser trap is generated and remains fixed in our configuration. Fig.2.13 (a) shows a schematic representation of the optical setup. The stiffness is calibrated through the power spectrum of a single particle's Brownian motion under optical tweezers [Neuman 2004]. On polystyrene microspheres of 3µm diameter, the trap stiffness is $1.8 \times 10^{-5} N.m^{-1}$ for the x-axis and $1.5 \times 10^{-5} N.m^{-1}$ for the y-axis (room temperature = 18°C).

A haptic interface (Omega, Force dimension), see Fig.2.13 (b), is coupled to a two dimensional motorized micromanipulator composed of two microstages for xy-

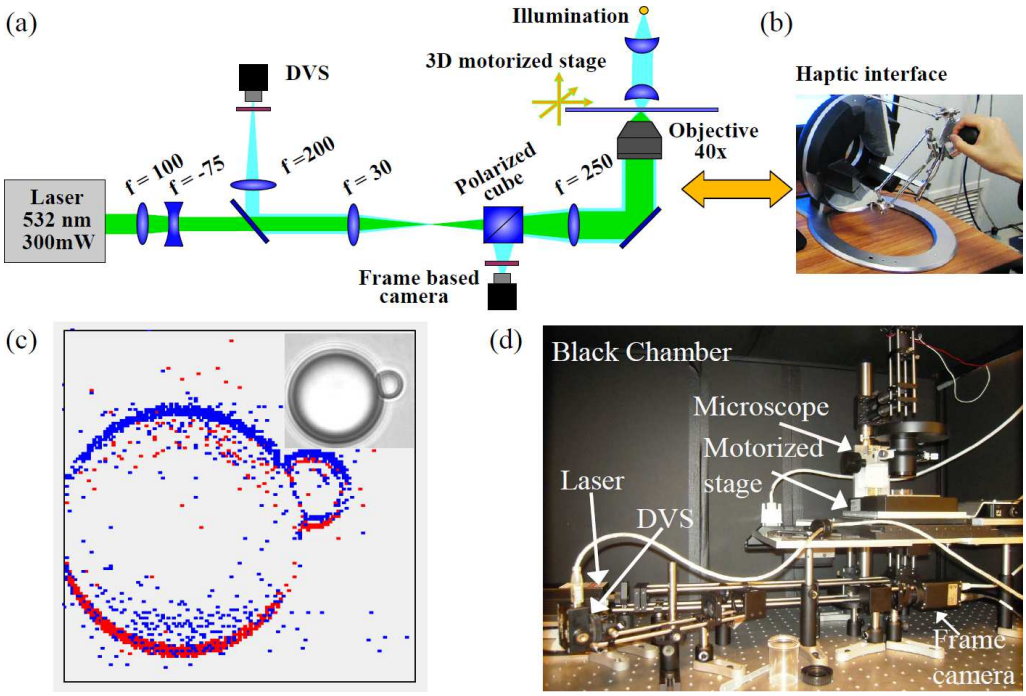


Figure 2.13: (a) A schematic representation of the optical system. (b) Haptic interface - the master device for teleoperation. (c) An accumulation map of events generated by DVS. Red and blue dots represent events of different polarities. The inset shows the correspondent scene in a frame based image. (d) A photo of the slave system of optical tweezers.

axes (PI, M126.CG). The haptic interface is coupled directly to the motorized stage with constant homothetic factors. The positions of the haptic device are multiplied by a factor of 4×10^{-4} to control the microstage and the measured force is magnified by a factor of 2.0×10^{11} to allow the force to be displayed to the operator (see Table 2.2 for parameters summary). Photos of the master system and the slave system are shown in Fig.2.13 (b) and (d), respectively.

Two types of cameras are mounted in our setup, a conventional CMOS camera (Basler 659×494 pixels) which allows visual perception for operators and the newly mounted DVS. Fig.2.13 (c) compares an event accumulation map of a bead probe touching a large sphere to the same scene captured by the frame based camera. Dots of different colors represent events of different polarities.

The implemented event processing algorithm is the Continuous Hough transform. Compared with other micro-particle tracking algorithms, e.g. centroid or cross-correlation, event-based Hough is able to lower the computational cost and meanwhile achieve better robustness.

Illumination conditions and the event generation threshold should be carefully configured in order to achieve the optimal bandwidth and the needed event rate. During the optical tweezers' stiffness calibration, the event rate lies around 30k

events per second (ev/s). In the scenario where the probe touches another object, the typical mean event rate is 50k ev/s. Statistically, a proportion of 30% of total events belongs to the movement of the probe and is used to update the circle position.

To compare with the DVS detection results, images are recorded from the frame based camera during manipulation. Hough transform is available as a built-in function in openCV computer vision library and is directly applied for performance comparison. In what follows the DVS processing is performed in real-time while images are processed off-line just for comparison purpose.

All system control, visual processing and data recordings use a single desktop PC with an Intel Xeon duo-core CPU running at 2.93GHz.

Trap stiffness	X-axis	$1.8 \times 10^{-5} N.m^{-1}$
	Y-axis	$1.5 \times 10^{-5} N.m^{-1}$
Scaling gains	Position	4×10^{-4}
	Force	2.0×10^{11}
Frame based sensor	Frame rate	30 fps
Event based sensor	Time stamp	1 µs
Haptic loop	Sampling rate	1 kHz

Table 2.2: Summary of important parameters

2.4.3 First Experiments on Force Sensing in the Microworld

In the search of instrumentation for cell characterization or organelle exploration, biologists require flexible and non-destructive tools to sense the microscopic interaction forces. An important task that has not yet been accomplished with optical tweezers is to trap a bead to touch an arbitrary target and then feedback the forces. The proposed experiments adapt to this emerging necessity, where a microworld touching and exploration is conducted. The probe used is a 3µm diameter polystyrene microsphere trapped by the optical tweezers. A big microsphere of diameter 10 µm is chosen as the target object, which can be viewed as a cell, a typical element in biological research.

2.4.3.1 Object Touching

The experiments will be conducted in two parts for two types of applications. The first one is the "come-into-contact". The probe is initially in a free state and then it starts to come into contact with a target object. The "come-into-contact" experiment is conducted to demonstrate the possibility of using the trapped bead as a probe to feel the forces during object contact. It initiates the basic utility that the probe can be used to sense the stiffness of an object.

Fig.2.14 side view schema illustrates the detail of this procedure. The user displaces the optical tweezers to drive the probe towards the target. Upon contact, the position of the laser spot is closer to the target than the center of the probe. This

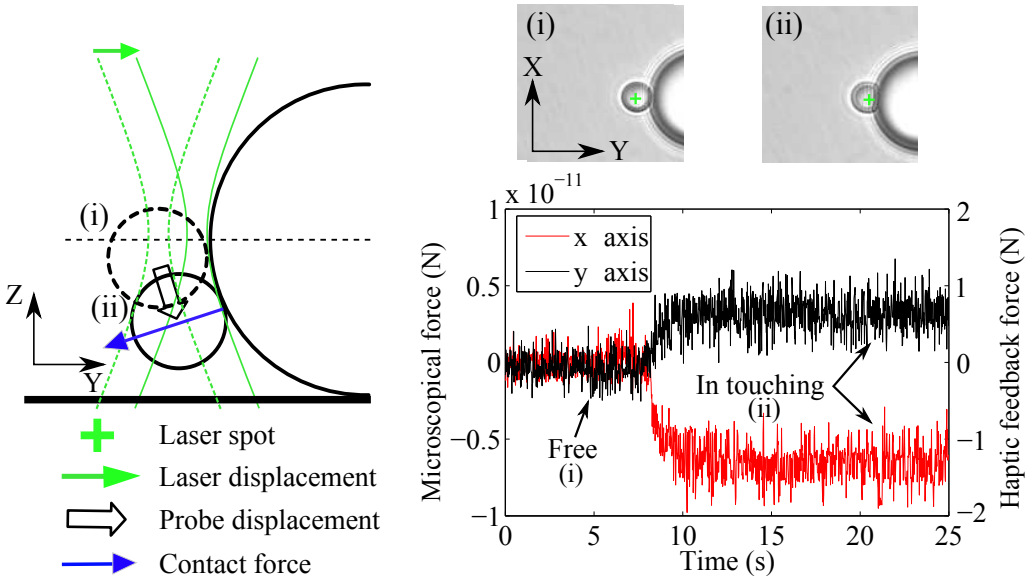


Figure 2.14: The probe "coming into contact" with the target object. Left image is a side view schematic representation of the procedure. The probe locations (i) and (ii) correspond to the inset images (i) and (ii). Insets (i) and (ii) show the moment before and while touching, respectively. Right below shows the microscopical and haptic forces measured by the probe during the "come-into-contact" experiment. The microscopical forces along the xy-axes of the probe have standard deviations of 0.83, 0.86 pN in free state and 1.17, 1.13 pN during contact.

deviated distance can then be applied to calculate the contact force. In our work, the spring force model is employed under the assumption that the optical tweezers remain in the linear region. Other complex force models taking into account more factors can also be envisioned. Since the probe is slightly below the medium height of the big sphere, the combinational force will drive the probe downwards thus the probe appears defocused, see inset (ii) of Fig.2.14. Probe positions (i) and (ii) in the side view of Fig.2.14 correspond to the top view inset photo (i) and (ii). The force feedback is two dimensional at the moment, which equals to the real force projected onto the focal plane.

Fig.2.14 right below illustrates the microscopical and the haptic forces detected by the vision sensor during the first part of the experiment. As expected, the force values start from around zero in free state and then change to other constant values when in contact. The microscopical forces along the xy-axes of the probe have standard deviations of 0.83, 0.86 pN in free state and 1.17, 1.13 pN during contact. Larger deviations appear when the probe keeps in touching. This implies a few instabilities of the visual detection algorithm in the contact situation. During contact, a small proportion of the votes in Hough space may be contributed by events generated from the contour of the target other than that of the probe, thus the detection precision is transiently reduced. This phenomenon is more obvious

when the decay period is tuned to be smaller. Longer decay periods can eliminate the artificial vibrations but will reduce the total bandwidth. We choose a 1 ms decay period as a compromise.

2.4.3.2 Surface Exploration

The second part of the experiment is to lead the probe to "walk around" a "cell". It allows applications such as scanning the surface topology of a biological element. In our experiment, we choose to use the probe to walk around the surface of a 10 μm microsphere for a first demonstration of proof-of-concept. There are two main reasons for selecting this task. Firstly, the exploration of a non-planar surface of a big microsphere is extremely difficult without force feedback. The user finds it hard to maintain constant contact with the circular surface thus the probe trajectory deviates easily. The trapped object escapes due to unstable equilibrium positions as shown in Fig.2.1. The combinational force may push the probe out of the optical trap due to weak trap stiffness on the z axis when contact happens above the medium height of the bead. Reliable force feedback can thus assist users to efficiently tune the interactions on the probe, to conduct it to walk around the shape and to prevent it from losing the trap. The second difficulty is from the point of view of vision. The color of the target object is the same as the "probe", which adds more complexities to visual processing than other state-of-the-art work.

Fig.2.15 shows the probe trajectory and the contact forces during surface exploration. The green solid curve is the laser spot trajectory and the black dots are the estimated probe positions, which both appear as a circular shape. During the task, force feedback assists the user in maintaining the probe consistently on the surface of the target. As shown in Fig.2.15, the force directions (blue arrows) remain the same as the radial directions of the large microsphere, which is an expected result. As demonstrated, the difficulty of the task is that the directions of the contact forces are continuously changing during manipulation. The forces are at the order of ten pico-Newton and are sent back to the user at the sampling rate of 1 kHz.

Without force feedback, it is difficult for the user to maintain a constant contact on the target surface. The sphere may escape easily or get stuck to the surface if the user pushes the probe too hard. By using the defocusing hint shown in Fig.2.14 inset (ii), experienced users may still perform this task. However, the concentration on the focusing status is laborious and tedious, thus impractical for long time manipulation. By enabling force feedback, users fulfil the presented task less laboriously, faster and more repetitively than without force. The probe escaping or surface sticking are also far less frequent.

2.4.4 A Comparison of Frame Based and Event Based Vision in Micromanipulation

Event based and frame based vision are completely different visual processing frameworks. Due to its redundancy elimination mechanism, the event based sensor and

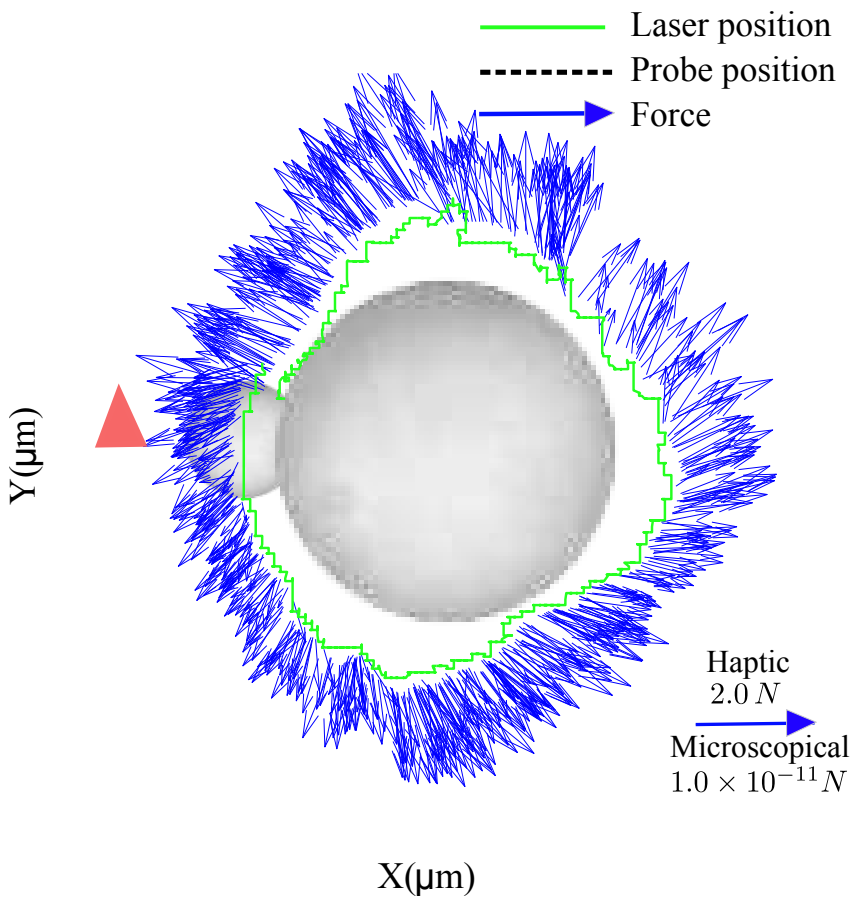


Figure 2.15: The positions and forces when the probe "walks around" the surface of the big sphere. The probe trajectory is depicted in black dots, the laser position in green line and forces in blue arrows. For clarity, the force arrows are subsampled to one thirty.

algorithms are able to achieve the same processing quality with respect to the frame based correspondences while providing higher detection bandwidth thus encourage us to exploit more complex algorithms. In order to allow comparisons, the power spectrum of the Brownian motion of the probe trapped by optical tweezers will be compared using both vision sensors. The power spectrum is one of the standard ways to calibrate the stiffness of the optical tweezers. The calibration thus requires the vision system to possess a high bandwidth far beyond the cut-off frequency of the power spectrum.

Fig.2.16 compares the results of the power spectra of the Brownian motion of the probe using both cameras. Fig.2.16 (light color) shows the spectrum using the frame camera running at 460 fps. A region of interest of 70×70 pixels is selected and the exposure time is manually diminished in order to achieve this high frame rate. The cut-off frequency is determined to be 60 Hz.

Fig.2.16 (dark color) shows the spectrum using DVS with the event based Hough

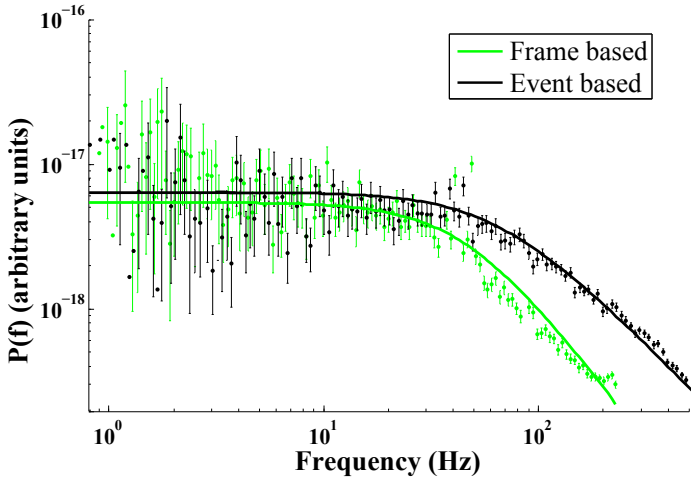


Figure 2.16: A comparison of the power spectra of Brownian motion of the probe detected by both the frame based camera at 460 fps (light color), and the event based vision sensor with equivalent frame rate 1k fps (dark color). The dots are raw data value with the bars their standard deviations. The solid curves are the fitted power spectra.

of 1 ms decay period, thus corresponding to 1 kHz equivalent frame rate. The bandwidth of the tweezers is estimated to be 80 Hz, slightly higher than that estimated by the frame based camera. This is because the frame rate of the conventional camera is not *far beyond* the cut off frequency of the optical tweezers, therefore high frequency signals are attenuated.

Special care should be paid in calculating the power spectrum using DVS raw data. Due to DVS's asynchronous timing and the Fourier transform implying a fixed interval sampling, the result will be flawed if the raw data is used directly. According to DVS's timestamp, the high rate asynchronous data, e.g. 30kev/s can be resampled to form a regularly sub-sampled sequence, e.g. 1 kHz. Our power spectrum is calculated from the resampled data. However, the resampled data is still hard to equally distribute exactly in time. The influence of the asynchronous data on signal processing will be out of the subject of the thesis and could be an interesting research topic.

Among the frame based particle tracking algorithms, Centroid or Histogram are the most frequently used. They are impossible to apply alone in a probe-target contact scenario because large bias will be produced when the calculation region contains obstacles, see Fig.2.4. Moreover, the similarity of light intensity of the probe and the target will further degrade the detection quality if non-robust image processing algorithms are used.

Frame based Hough circle detection algorithm is immune to obstacles. OpenCV's implementation of Hough transform provides real-time performance under the criteria of conventional computer vision in the order of tens of frames per second. It hence has never been successfully applied for high speed micro-particle tracking.

Moreover, luminance and focus play an extremely important role in edge detection, which is the prerequisite of the frame based Hough transform. Slight differences in optical configurations may produce large detection uncertainty.

It is thus important that DVS serves as a robust hardware edge detector. By employing the edge information directly, the presented method has been applied successfully to track a probe in contact with an object of arbitrary form and color, which shows the advantage of our method facing obstacles and environmental noise.

We are not meant to completely abandon the frame based approach. The key advantage of using event based vision is that traditionally complex algorithms under frame based framework can be easily assimilated to achieve high speed processing. By employing the event based method, the effective position refreshing rate of the probe attains 15 ± 5kHz during the "walk around" experiment while all the system controls and haptic coupling are running. The processing speed is much higher than the desired haptic feedback frequency of 1 kHz. Of course, static information is not detected by DVS, and if required, the frame based processing is still a perfect complementary, as will be shown in Chapter **??**.

Conclusions

In this chapter, a high speed circle detection method – the Continuous Hough Transform based on an unconventional event based vision sensor – DVS, is presented. The method is shown to be more robust than existing high speed vision detection methods in micromanipulation and is able to achieve the same detection quality as a far more complex frame based algorithm that runs one order slower. The tracking capability is validated and the performance is analysed by tracking multiple microspheres moved by microstage at 30 kHz and by high dynamic Brownian motion detection.

Haptic force feedback assists operators in manipulating objects, keeping a reasonable contact force, blocking for dangerous manipulation, etc. We have coupled the haptic device to the optical tweezers. The system presented in Section 2.4 demonstrates the possibility of high bandwidth reliable pico-newton level force sensation by using event based vision sensor and initiates a broad range of microscopical applications. The ideal haptic feedback sampling frequency of 1 kHz is successfully achieved. Complex tasks of manipulating the probe for object touching and non-planar surface exploration have been demonstrated for the first time.

However, the force feedback is still limited to the xy plane. In order to build a complete and dexterous system to fully exploit the capability of optical tweezers, the movement of the probe and the force feedback should be accomplished in the whole 3D space. New event based position detection methods should be designed to perform 3D tracking. The issue of out of focal plane movement appears, thus adding extra difficulties for tracking. Moreover, the current experimental platform must be modified to provide 3D functionality. Multiple optical traps with parallel force measurement can also be envisioned in the new setup. These interesting materials

will be the content of Chapter 3.

Chapter 3

Asynchronous Event Based 3D Microsphere Tracking

Contents

In the previous chapter, a haptic feedback teleoperation system of optical tweezers is successfully developed to show the possibility of using the trapped bead as a tool for force sensing during contact. The event based Hough transform is able to track the center position of a microsphere with known radius at high speed and reliable tactile sensation is achieved. However, the force feedback remains two dimensional, namely limited to the xy plane. An ultimate system that manipulates and at the same time feeds back forces in three dimensional space and functions in parallel with multi-trap optical tweezers will surely benefit various types of applications. The force probe or probes can then be used to explore freely in the whole workspace, thus the user can sense the forms and stiffness of biological and medical samples in the real-world environment regardless of the focal plane.

This chapter shows how to build such a system. The existing system in Chapter 2 is no longer adapted. More actuation units must be mounted to enable 3D displacement and simultaneous control of multiple traps.

More importantly, to enable 3D force feedback, high speed position detection must be accordingly performed in 3D space. Due to the expanded capability and the blurring effect caused by the defocus of the scene frequently encountered in 3D movement, a new algorithm must be designed. How to tackle defocus? Is it still possible to keep using the event based vision? What is the best hint for the detection of position in the z axis using a monocular vision system? These questions should be answered. In this chapter, various algorithms, either extended from the previous Continuous Hough transform or newly developed from scratch, have been tested. Finally, a robust circle fitting method is chosen to accomplish the detection in the whole 3D space and to permit multi-trap tracking.

In this chapter, it will be shown that for the first time, real 3D exploration tasks can be performed in random liquid conditions in a repeatable manner at high operating speed by an untrained operator. A patent has been deposed for this invention.

3.1 3D Sphere Tracking Methods

Axial positions can be determined by stereoscopic structured illumination [Bowman 2010], by holograpic techniques [Cheong 2010] and by defocus hint [Zhang 2008].

Most of the existing microscopes observe samples either from above or below. Side views are just rare. An easy-to-use teleoperation system should thus be easily integrated into existing microscopes that provide only monocular and planar views. In other words, by simply installing some additional setup, e.g. haptic device, and associated tracking software, the haptic feedback micromanipulation system should then be readily functional. Embedding that in mind, we aim to design such a robotic tool that can easily supplement the publicly available microscopic systems that finally assists operators to perform delicate micromanipulation or microassembly tasks. As a result, it would be essential to find out clues of displacements of microspheres in the z axis from the xy view alone, since the popularity of stereoscopic or holograpic microscopes is very limited. Interestingly, objects focus or defocus with respect to the focal plane, which constitutes valuable information to exploit the third dimensional location.

3.1.1 Defocus

Defocusing is a double-edged sword. On one side, it provides a useful hint that permits the detection of the depth information and allows monocular vision to possibly estimate three dimensional position. On the other side, defocused images bring special difficulties in image processing. It requires more robust algorithms to tackle object forms appearing in different depths. Fig.3.1 shows a sphere in the same position of the xy plane but varying in z. Diffraction generates the spherical radial profile that exhibits several concentric circles and variable light intensity. From sphere center to the outermost edge, intensities oscillate, which is exactly

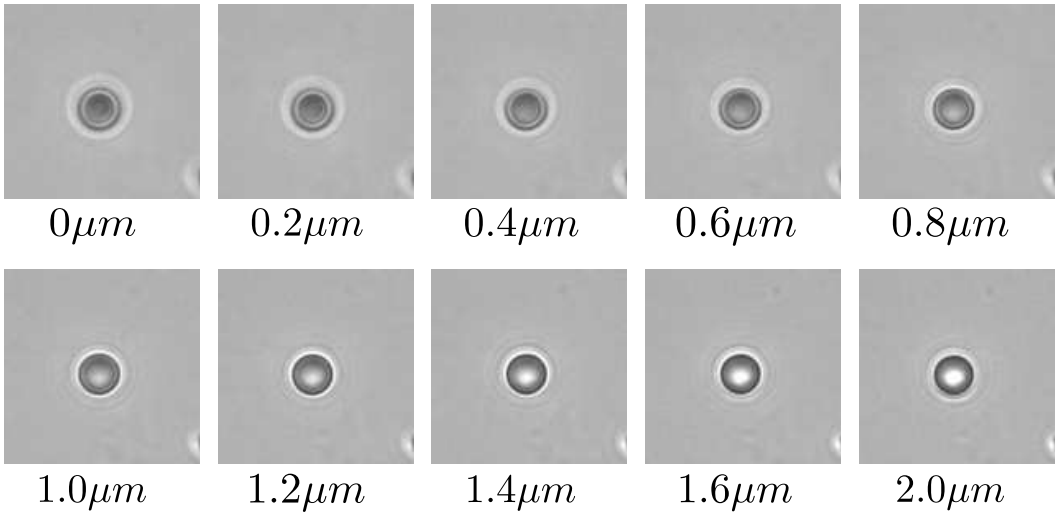


Figure 3.1: An sphere of diameter 3 µm stays in the same location of the xy plane but varies in z. Diffraction generates the spherical radial profile that exhibits oscillating light intensities.

the information that has been used to track spheres trapped by magnetic tweezers [Gosse 2002]. The radial profile is used after a calibration step, in which each profile at a given z location is recorded, see Fig.3.2. Then the z position detection consists in matching the current profile to the recorded one by minimizing least square errors. Interpolation between recorded profiles is then performed to achieve higher resolution. The interpolation relies upon the wave aspects of the diffraction rings. It shows that the phase shift between successive profiles is fairly linear with their distance from the focal plane. Unfortunately, this method has a major limit because it depends on the precise defocus profile. Any contact scenario will degrade the radial symmetry so that the observed profile will be different from the recorded one in the calibration image. Moreover, different optical setups will produce fairly different diffraction profiles so that the calibration step should be preformed once again even with a slight change of optical conditions.

3.1.2 Intensity Average on Frame Based Images

Since the xy position has already been detected using DVS, one possible way to perform 3D tracking is to combine the DVS with a frame based camera. Using this method, it remains to perform a classical image processing to retrieve depth information from defocus, certainly with much lower frequency than DVS, e.g. $120Hz$. The processing could be as simple as computing the average gray level in the surface area of the tracked probe, such as:

$$\overline{I}_S = \frac{\iint_S I(x,y)dxdy}{\iint_S dxdy} \tag{3.1}$$

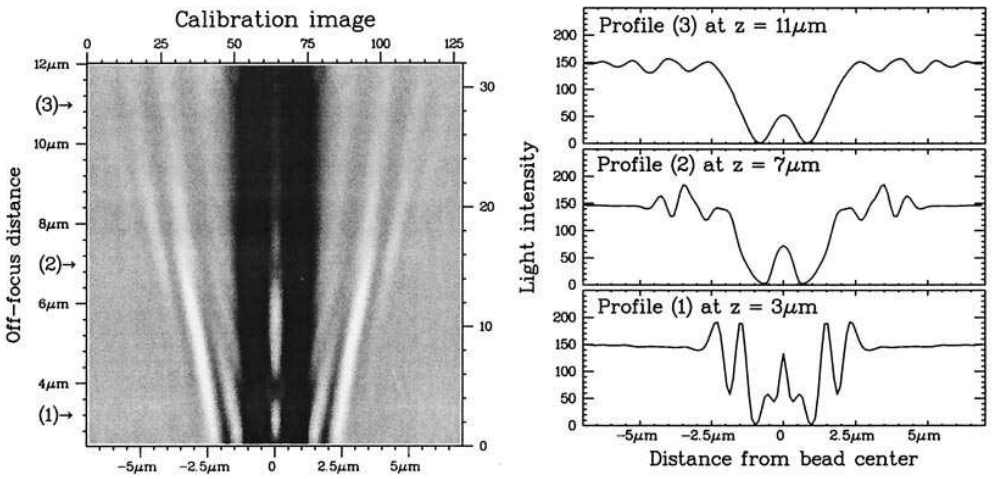


Figure 3.2: (left) Calibration image of a 4.5 μm bead laying on a glass surface. Each line is obtained by measuring the radial profile of the object at a specific z position of the microscope objective. (right) Intensity profile corresponding to 3, 7, and 11 μm in z axis. Image courtesy of [Gosse 2002].

where $I(x, y)$ is the image intensity in pixel location (x, y) and S is the surface area of the microsphere, the location of which can be obtained from the result of continuous Hough Transform, see section 2.2. Originally represented in DVS coordinates, a calibration step is required to transfer the center location of the probe from the coordinate system of DVS to that of the frame based camera. This relation can be represented by Homography, refer to section **??**. Then each location of z can be related to a specific gray level $\overline{I}_S$, thus providing the possibility of useful haptic assistance in the third dimension.

However, this method is not robust during contact. Remind that both DVS and frame based camera are constantly centred on the probe, so when the probe moves along the vertical axis, objects other than the probe will change significantly their appearance due to defocus in the z axis.

An experiment has been conducted to make the correspondence between the gray level $\overline{I}_S$ and the location of z. During the test, the xy plane movement is disabled and only the z axis is active. The user tries to maintain the probe in contact with a 10 μm microsphere and meanwhile to move the probe vertically. The average intensity on the surface of the probe is calculated. Fig.3.3 shows a non-linear relation between the z displacement of the probe with respect to the medium line of the 10 μm target sphere and the average gray level intensity $\overline{I}_S$. A second order polynomial is used to approximate this relation, which is shown to be $\overline{I}_S = z_0 \pm 1.6z^2$. In Fig.3.3, the branch above (the '+' sign part of the fitted equation) shows the relation when displacements of the probe in z is upwards with respect to the medium line of the target, while the branch below (the '-' sign part) signifies the displacements downwards. The impact of third party objects on the detection correctness is demonstrated. Fig.3.3 insets shows diffraction ring interferences and

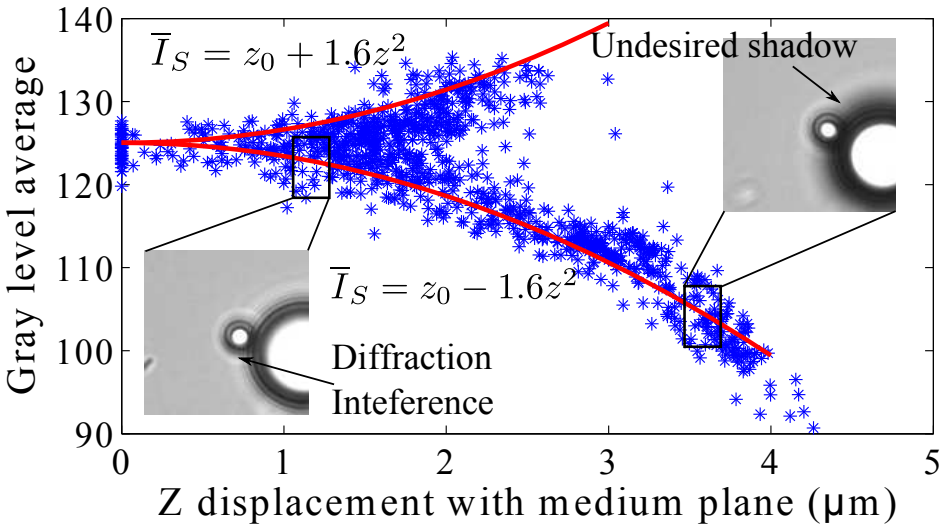


Figure 3.3: A non-linear relation exists between the z displacement of probe with respect to the medium line of the 10 µm sphere and the average gray level intensity $\overline{I}_S$. A second order polynomial is used to approximate this relation. The branch above (the '+' sign part of the fitted equation) shows the relation when the displacement of the probe in z is upwards with respect to the medium line of the target, while the branch below (the '-' sign part) signify the displacement downwards. The insets shows diffraction ring interferences and undesired shadow projection on the probe.

undesired shadow projection that make the detection result very ambiguous. The average intensity $\overline{I}_S$ thus not only depends on the relative distance between the laser and the probe center, but also on the probe-target relative position, which is strictly undesired.

The quadratic fitted relation $\overline{I}_S = z_0 \pm 1.6z^2$ does contain the real force generated by displacements of the probe with respect to the laser center. However, the impact of diffraction or shade laying on the probe produces important noise, the proportion of which is difficult to quantify. Therefore, this method provides some combinational information on the depth, contact status and defocusing, and thus is still a very useful hint for haptic virtual assistance. However, the method is neither robust for precise position measurement nor fast enough. In consequence, the combination of event based and frame based cameras should be excluded from the final solution.

3.1.3 Polarity Integration

It is ideal to find the z axis position by using DVS data alone without the participation of frame images. This is not only a more compact solution but also overcomes the difficulties often encountered in the frame based method.

Up to now, the polarity of events has not yet been taken into account. Polarities

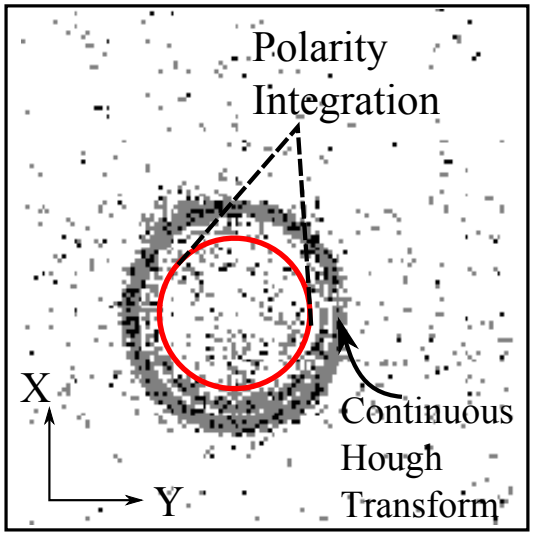


Figure 3.4: Polarities supply valuable information indicating directions of movements. The integral of events of different polarities is calculated within the red circle region, while the events on the edges are used by continuous Hough transform.

supply valuable information indicating directions of movements. Intuitively, when the probe moves upwards, the events generated in the central region of the sphere will be dominated by $+1$ polarity, while descending process will be dominated by events of -1 polarity, see Fig.3.4. Fig.3.5(a) shows the mean event rate during the sinusoidal movement of the sphere in z axis. The sign of the polarity is taken into account in computing the rate. To clarify, a $+1$ event and a -1 event together will produce zero mean event rate. It shows clearly that the dominant event polarity can definitively reflect the movement of the object on the vertical axis. Therefore, by integrating events of different polarities, one should be able to acquire the depth information in z.

The approach is thus to sum up the events, in the process of which polarities provide the signs for integral computation. Using the definition of events in Equation 1.4, the calculation can be written as:

$$\int_0^{+\infty} \iint_S e(\mathbf{p}, t) d\mathbf{p} dt \tag{3.2}$$

in which S is the central region of the circle, shown in Fig.3.4 the area within the red circle. Note that events outside of the red circle are used by Hough transform thus are not considered in this procedure.

Unfortunately, the polarity method suffers from one major problem. Even if the probe moves exactly the same distance in opposite directions, the number of ON events and OFF events will never be identical. This explains why the integral of polarity of events has always an approximate sinus form, but as time passes, the

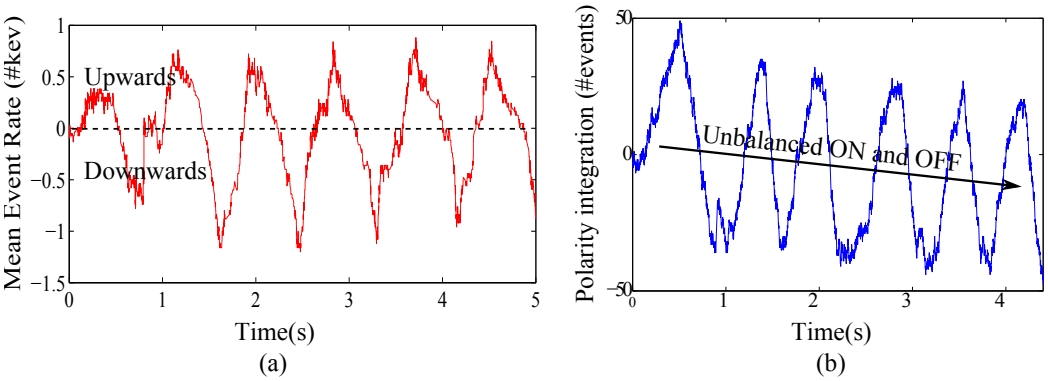


Figure 3.5: DVS observes the probe doing sinusoidal movement in the z axis. (a) The mean event generation rate in the region inside the circle. (b)The integral of events of different polarities. Unbalanced ON and OFF events are observed.

mean value deviates towards a certain direction that depends on the DVS sensitivity of ON and OFF. This shift in average is a typical problem using integral associated methods that present accumulated errors, see Fig.3.5 (b).

In summary, this method is interesting to provide a general idea of whether the probe is ascending or descending, or possibly with resolutions of several steps, namely only several discrete steps upwards or downwards to be quantified. It requires quasi-zero computational resources and is independent of xy position detection. However, it is incapable of achieving a continuous and precise estimation of the z axis position. A more adaptable method should be studied.

3.1.4 Extension of Continuous Hough Transform

During defocusing, a useful hint is the slight change of circle radius, which appears to be fairly linear with a small z displacement within $\pm 0.5\mu$m from the laser focus. Based on Chapter 2, the 2D continuous Hough transform can be extended to 3D space. In this case, the triplet (x_c, y_c, R) will be estimated instead of the pair (x_c, y_c). The supplementary dimension R can then be converted to the depth information after a calibration stage.

For 3D Hough, an event will vote a cone instead of a circle in Hough space. This is a time consuming procedure and is often avoided by separating it into two successive lower dimensional procedures [Ioannou 1999]. This type of methods, however, requires the use of complete image edges thus is not applicable to event based data. We thus choose to augment directly one dimension in Hough space. The process will certainly be slowed down and events may be lost. The difficulty can be partially overcome by increasing the sampling angle of voted circles to every 5 degrees, and by restricting the radii from possible minimum value to maximum value, typically not exceeding 5 pixels.

The radius change is very slight, not as evident as the displacement in the xy

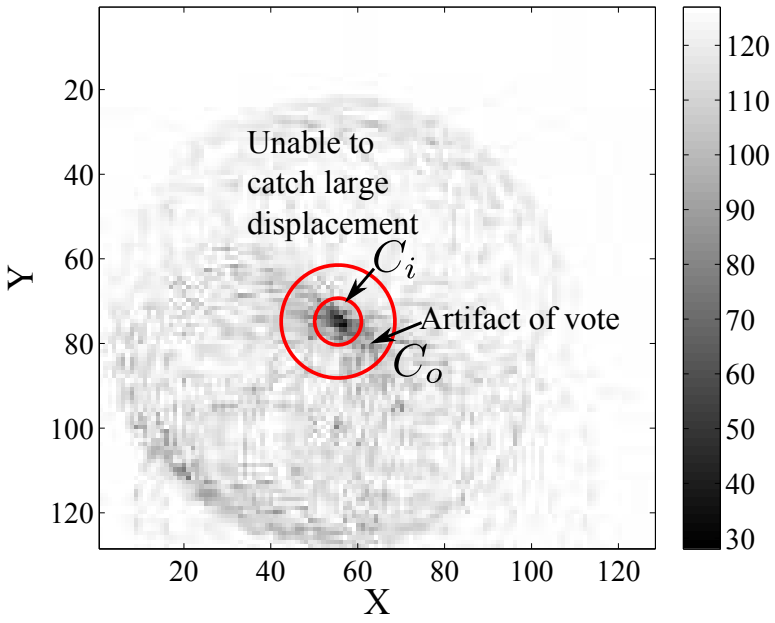


Figure 3.6: Hough space. If the centroid region is as small as C_i, the algorithm averages only in the peak region thus does not support large scale displacement. On the contrary as in region C_o, artefacts become influential due to improperly including edges of voted circles.

plane, so a small quantity of noise will have a strong impact on the detection result. Calculating the centroid in Hough space will generate artefacts, which constitutes a major source of detection noise, see Fig.3.6 the Hough space is shown in 2D for easy illustration. If the centroid region is small, as in C_i, the algorithm averages only a limited region surrounding the peak thus prohibiting large scale displacement. On the contrary, if the centroid region is large as the region C_o, artefacts will be included into the process due to edges of voted circles. The inconvenience has not been much observed in Chapter 2 because the centroid region is limited to two dimensions and the sphere radius spans only 10 pixels. The problem is more evident in 3D computation, where the centroid tries to average the cones. Furthermore, lower processing speed makes fewer events be considered within 1 ms. The resulted small signal to noise ratio degrades the detection credibility.

In summary, 3D Hough transform is a working but not completely satisfying method. The noise level is high and the computation load is expensive. A much more suitable algorithm will be presented as follows.

3.1.5 Robust Circle Fitting

Pixel locations of events are intrinsically points. Numerous point cloud processing methods are thus inspiring [1]. In our case, a circle fitting method is suitable to match the events' location to a parametrized circle.

Let vector $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ be a vector of unknowns and consider the system of m equations $f(\mathbf{u}) = 0$. For parametric curve fitting, the goal is to look for:

$$min(\Sigma_{i=1}^{m} f_i(\mathbf{u})^2). \tag{3.3}$$

Specifically, the algebraic representation of a circle with its center at (x_c, y_c) and radius R can be written in the form of:

$$(x - x_c)^2 + (y - y_c)^2 = R^2. \tag{3.4}$$

To estimate the parameters of a circle, Equation (3.4) should be reformulated into a linear system in the form of $A\mathbf{u} = \mathbf{b}$.

Then, let us set $u_1 = -2x_c$, $u_2 = -2y_c$ and $u_3 = x_c^2 + y_c^2 - R^2$. Expanding the event location $\mathbf{p_i} = (x_i, y_i)$ into the linear system, one achieves:

$$\begin{pmatrix} x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots \\ x_m & y_m & 1 \end{pmatrix} \mathbf{u} = \begin{pmatrix} -(x_1^2 + y_1^2) \\ \vdots \\ -(x_m^2 + y_m^2) \end{pmatrix} \tag{3.5}$$

Equation (3.5) is usually over-ranked. A pseudo-inverse or singular value decomposition (SVD) can be employed to solve this equation in a least square sense. Finally, the circle center (x_c, y_c) can be simply obtained from $(-u_1/2, -u_2/2)$ and the radius R from $((x_c)^2 + (y_c)^2 - u_3)^{1/2}$.

This basic version of circle fitting is called "ordinary least square method (OLS)". A more complex modification allowing the implementation of robust fitting methods requires to slightly reformulate the above equations. Weights could be supplemented into the linear system that permits to adjust the contribution of each point to the minimization of the square sum of residuals, according to some criteria.

An OLS tries to minimize the following square errors of a linear system $A\mathbf{u} = \mathbf{b}$:

$$F = (A\mathbf{u} - \mathbf{b})^T (A\mathbf{u} - \mathbf{b}), \tag{3.6}$$

and its general solution is given as:

$$\mathbf{u} = (A^T A)^{-1} A^T \mathbf{b} \tag{3.7}$$

To allow weights, the target minimization function can be rewritten as:

$$F = (A\mathbf{u} - \mathbf{b})^T W^{-1} (A\mathbf{u} - \mathbf{b}), \tag{3.8}$$

[1] http://pointclouds.org/

therefore the associated general solution of the minimization problem is modified correspondingly:

$$\mathbf{u} = (A^T W^{-1} A)^{-1} A^T W^{-1} \mathbf{b} \tag{3.9}$$

Once a first guess of parameters are obtained, the above procedure is then repeated until an error tolerance is attained or the iteration times exceeds a threshold. This modified least square method is called "iterative reweighted least squares" in the literature, of which the derivation detail can refer to [Zhang 1997].

Numerous studies show that least squares are vulnerable to outliers. Even in the presence of only one bad datum far away from the ground truth, the whole estimations procedure may be completely perturbed. Robust methods have been proposed to overcome the sensitivity of least square to outliers [Huber 2009]. M-estimator is such a well recognized robust method for eliminating the impact of outliers by intelligently choosing the associated weights. In short, good data will have more dominant weights while bad data will logically have trivial weights. The influence of outliers can thus be significantly reduced.

The implementation of M-estimator can be written in the same way as the iterative reweighted least square. The only problem left to be addressed is how to compute the weight matrix W. Lots of functions can be used as the weight function. [Zhang 1997] gives a fairly complete summary and methods of how to choose among the functions according to the requirements of applications. For example, L_ν type represents a family of power functions. Let x denote the residual of points, namely the error between the observed and the fitted data. The weight function of L_ν type can be written as:

$$\omega(x) = |x|^{\nu-2} \tag{3.10}$$

Another useful function that has been recommended for almost all situations is the Huber function such as:

$$\omega(x) = \begin{cases} 1 & if\ |x| \leq k \\ k/|x| & if\ |x| > k \end{cases} \tag{3.11}$$

Fig.3.7 provides a visual representation of the L_ν and Huber weight function. As expected, the weight function has maximum value in the region around the origin and decreases away from zero in order to balance the influence of different data points. Logically, the robust method will consume much more time than the basic fitting method. The choice should be made in preference of speed or robustness. To resume, the robust circle fitting method can be written as in Algorithm 2.

3.1.6 Summary of Different Methods

Numerous methods have been developed to detect positions in the z axis. Using structured stereoscopic illumination or holographic techniques requires a major redesign of currently the most popular laboratory microscopes. The goal here is to provide an easy to use force feedback system based on the widely available microscope designs of general public. Therefore a third option is selected by using defocus

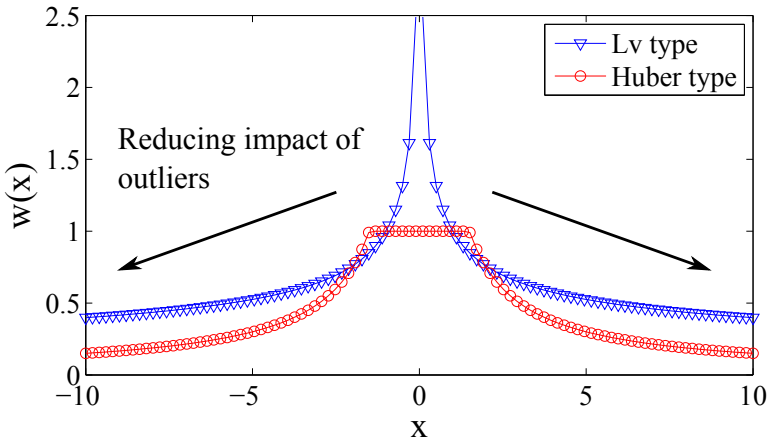


Figure 3.7: Weight functions of L_ν type with $\nu = 1.6$ and Huber type with $k = 1.5$, serving the role of reducing the impact of outlier data.

Algorithm 2 Robust Circle Fitting

1: **for** every 1 ms **do**
2: Compute the least square solution of $\mathbf{u}$ by solving Equation (3.5) or (3.9).
3: Transform the parameter vector $\mathbf{u}$ into real circle parameters (x_c, y_c, R).
4: Calculate the residues.
5: Update the weight function $w(x)$ according to Equation (3.10) or (3.11).
6: Repeat step 2-5 until residues are below the threshold or maximum iteration times are attained.

from a monocular vision system, as the hint to provide the sphere position in the third dimension.

Five methods have been presented in the previous section. Table 3.1 gives a succinct comparison of the advantages, the disadvantages and the complexity of each method. The (*) symbol indicates the relative complexity of the algorithm, with three stars representing the most complex case. It is shown clearly that the robust circle fitting method is currently the best choice.

3.2 3D Haptic Feedback Teleoperation of Optical Tweezers

By redesigning the position detection algorithm and reconstructing the optical system, force feedback can be a reliable scientific instrument to observe and interact with phenomena occurring at the micro-scale. A three-dimensional shape is explored with reliable tactile sensation. Difficulties, such as defocus of the probe, are anticipated by the vision process during the manipulation tasks.

Name (Complexity)	Capability	Properties	Type
Radial profile matching (**)	Sub-pixel resolution	Strictly symmetric; Illumination dependent	Frame
Intensity average (*)	Provide simple hints for virtual assistance	Strongly perturbed by shades	Frame
Polarity integration (*)	Several discrete steps	Very limited resolution	Event
3D Hough Transform + Centroid (***)	Sub-pixel resolution	High noise; Time consuming	Event
Robust circle fitting (**)	Sub-pixel resolution	Low noise; Fast	Event

Table 3.1: Comparison of the presented position detection methods for z axis. The complexity of the algorithm is relative, with (*) indicating the least time consuming. The column "type" lists whether the method is frame based or event based.

3.2.1 Configuration and Method

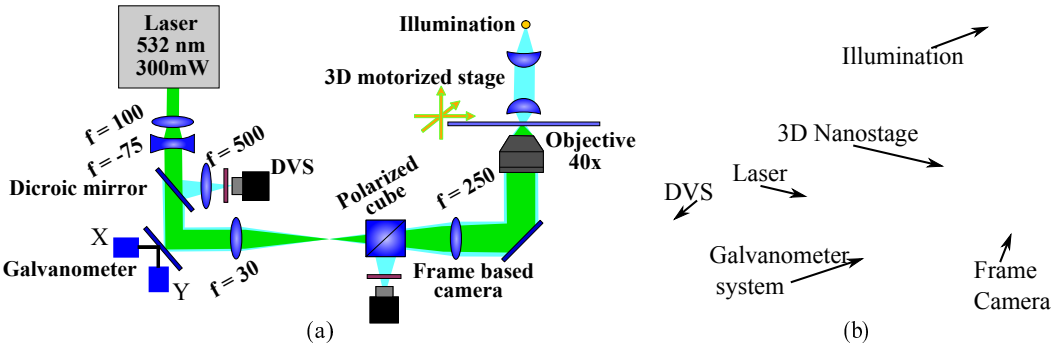


Figure 3.8: (Left) The new setup for the 3D haptic feedback multi-trap optical tweezers system. (Right) A photo of the system.

The experimental system contains part of the design as in Section 2.4: a laser source (532nm, 300mW), a high numerical aperture objective (40x, NA.1.3, Olympus), two micro stages for xy axis (M126.CG,PI) and an extra nanostage for z-axis (P528.ZCD, PI), see Fig3.8. A fast CMOS digital camera of 659×494 pixels (Basler scout) allows the visualization of the scene. The second camera is DVS for particle tracking. The galvanometer is an added device for time-shared multiple optical trapping, which will be presented in detail in later sections. The haptic device used is Omega (Force dimension). The scaling gains are 1.6×10^{-4} for the position and 1.8×10^{11} for the force magnification. No further coupling control has been necessary to obtain a stable and useful system.

The probes are polystyrene microspheres of 3 μm. They are treated with a surfactant (Tween-20) to prevent adhesion during contact.

Modifications have been made to ameliorate the optical quality of the setup. Remember that the more evident the change of radius, the higher the detection resolution. Compared to Fig.2.13, the optical path is changed to achieve smaller depth of field, resulting in ameliorated resolution in the z axis and also in the xy plane as a bonus.

According to geometric optics, the limit of the depth of field is influenced by several factors that can be resumed in the following formula:

$$Depth\ of\ field = \frac{Working\ distance}{1 \pm Blur\ spot * Iris * \frac{Working\ distance - Focal\ length}{Focal\ length^2}} \tag{3.12}$$

Depth of field thus relies upon three types of parameters: the blur spot (pixel size), the open of iris and the focal length. Pixel size and iris cannot be changed in the setup so we concentrate on modifying the focal length of the lens situated in front of DVS. It shows in Equation (3.12) that the longer the focal length, the smaller the depth of field, the relation of which is quadratic. A lens of 500 mm focal length is employed, in such case the radius of sphere will cover 33-34 pixels.

The optical trap stiffness has been calibrated on the microprobe with $1.6 \times 10^{-5} N/m$ for x-axis and $1.4 \times 10^{-5} N/m$ for y-axis using the power spectrum method. It should be clarified that the forces feedback to users on z are not necessarily a magnification of the real microscopic forces due to the lack of an effective force model on the axial direction of the optical trap. The z axis stiffness has not been calibrated due to this reason.

A monocular DVS is used for both xy planar force measurement and z axis position detection. During the tracking process, a distance check is performed to eliminate events generated far away from the circle edge. Afterwards the circle fitting algorithm is applied to estimate the center and the radius of the probe at the speed of 1 kHz. In preference of speed and to leave some margin for multi-trap, the basic version of the least square circle fitting algorithm is employed. It shows that elimination of events by distance checking is sufficient to guarantee robustness in the application.

The status of beads, either trapped or lost, is constantly checked by computing the mean event rate. If the event rate approaches near zero, it indicates that no bead is trapped so the force should be instantly turned off for safety.

Contact and non-contact modes are permitted in the present system. When the forces (or the sphere-laser distance) start to increase above a threshold, e.g. 1.75 pixels, the system enters the contact mode, and returns to non-contact mode when the distance is below 1 pixel. The hysteresis is purposely set to avoid switching back and forth frequently between the two modes. The threshold is chosen according to experience, guaranteed to be higher than the force value of Brownian motion. The interest of mode setting allows the possibility of special treatment during probe contact, such as adding blocking force or performing a filtering of force to strengthen stability. It is important to note, however, that mode change is only an add-on

capability of the system. The haptic feedback is totally functional with this mode turned off.

3.2.2 Z Axis Force Feedback

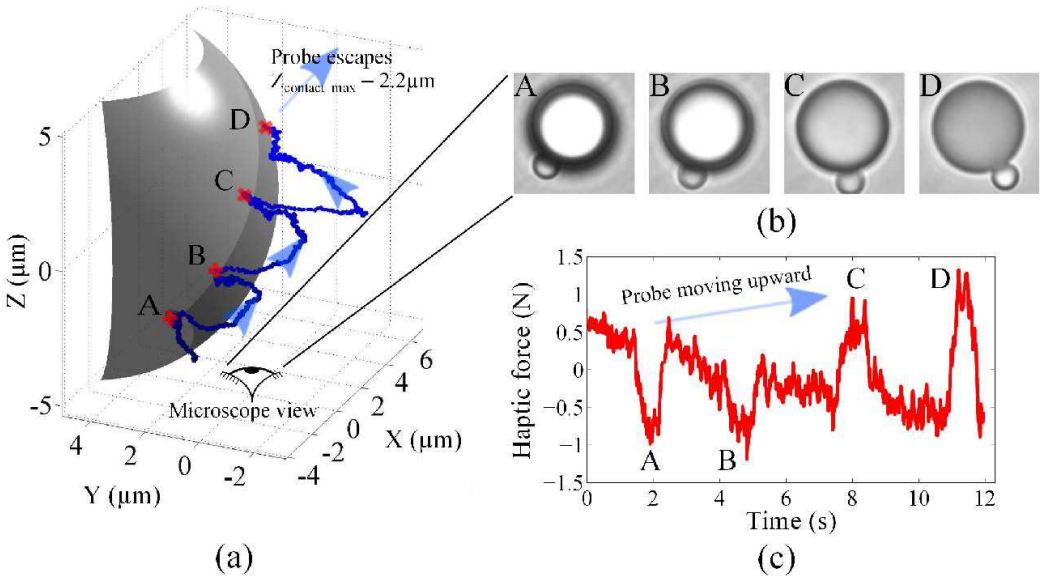


Figure 3.9: (a) The operator touches a spherical surface in different 3D points A,B,C and D. The probe trajectory is in blue. (b) Images show the views that an operator observes in points A,B,C and D. (c) shows z axis forces, where peaks are the moments of the probe-target contact. The force values change with depths of the optical trap.

Results of 3D experiment, more specifically, the forces on the z axis are shown in Fig.3.9. The operator tries to touch a spherical surface in different 3D locations (points A,B,C and D in Fig.3.9) from bottom to top. The probe trajectory is shown in blue curve in Fig.3.9 (a). The probe finally escapes at 2.2µm above the medium line of the surface. This experiment is especially designed to test the force quality in z axis. The image inset (b) provides views that an operator observes in points A,B,C and D. Images appear very differently in different z locations where the frame based methods will surely encounter difficulties. The consequent z axis forces are shown in (c), where peaks are the moments of the probe-target contact. As expected, two peaks point downwards (points A and B) and the other two point upwards (points C and D). An interesting phenomenon observed is the variation of force values with the depths of the optical trap. Unlike the fact that the forces are indifferent of the optical trap location in xy plane, this phenomenon provides experimental evidence that a simple spring model is inadequate for correct z-axis force measurement. This is however helpful in practice because haptic forces thus have the tendency of pulling the probe back to the focal plane.

Experiments similar to Section 2.4.3.2 can now be conducted easily in 3D space. Under the assistance of the haptic feedback, following the contour of a 10 μm microsphere can be completed within a short period not more than 10 seconds.

3.3 Haptic Feedback on Multi-trap Optical Tweezers

Multiple optical trapping will significantly enhance the capability of micromanipulation. A laser can be separated into multiple beams to simultaneously trap objects in different locations, thus allowing parallel micromanipulation. Coupled with parallel haptic device, the system will be able to control multiple microspheres simultaneously to form a "microhand", in which each probe acts like an independent finger of a human hand. The ultra-flexibility and dexterity of such system will surely open a greater range of microscale applications that one could ever imagine.

3.3.1 Time Multiplexing Multi-trapping by Galvanometer

In order to achieve multiple object trapping, space multiplexing or time multiplexing methods are possible. Spatial Light Modulator (SLM) is currently a standard way to perform space sharing by generating holograms. SLM is actually a dynamic computer-controlled diffractive optical element that allows maximally a few hundred traps running simultaneously. Its phase modulation capability by generating holograms is highly efficient in terms of laser energy distribution.

Figure 3.10: Laser actuation: (left) Passive mode where the laser is fixed while the platform moves; (right) Active mode where the laser moves. The green crosses represent the two laser trap centers.

A bilateral teleoperation system of holographic optical tweezers (HOT) has been developed [Onda 2012a]. The control speed of HOT has been strictly limited to the calculation speed of holograms. Recent progress in parallel computing made it possible to generate holograms in real-time by graphics processing units (GPU). The haptic loop rate reported in the design of [Onda 2012a] is 100 Hz. The software technique often employed in such system is a combination of Labview and OpenGL. OpenGL running on a parallel GPU processor is capable of generating holograms at hundreds of hertz. Labview is used for both graphic user interfaces and particle tracking, in which the center of mass and a symmetric based method are provided.

Another interesting demonstration is the "iTweezers". It uses a touch screen iPad to control the positions of about ten microspheres by sliding fingers on the screen.

Similar systems are reported by both [Onda 2012b] and [Gibson 2012] in the same year. The "iTweezers" system does not actually provide force feedback because the iPad is simply a position control device. In fact, it lays another important reason why the force feedback is difficult, which relates to the laser actuation method and the calibration of the laser center. In optical tweezers system, two types of laser actuations are possible. If the laser beam is fixed in the camera view and only the micro/nanostage moves, it is called "passive" actuation, see Fig.3.10 (left). In this setup, once the camera is centered on the laser, the whole camera resolution will be used to image the close neighbourhood of the trapped bead so any slight deviation of the bead from the laser center will be captured. However, by using laser actuators, such as SLM or Acousto Optic Deflector (AOD) [Ruh 2011], the laser beam becomes "active" and may no longer be fixed in the camera view, see Fig.3.10 (right). In the scene, the trapped object travels a large range with the movement of laser spot. The bead-laser distance is thus difficult to calculate. In summary, although it is able to generate many traps and is proven to be very energy efficient, successful haptic feedback on holographic optical tweezers is rare due to the difficulties in the laser center calibration and the resulting lower force sensing resolution.

Time multiplexing by a galvanometer scanning mirror system is a solution to circumvent the problem. In this section, the goal is to realize parallel force feedback and to provide a first proof-of-concept by generating double traps.

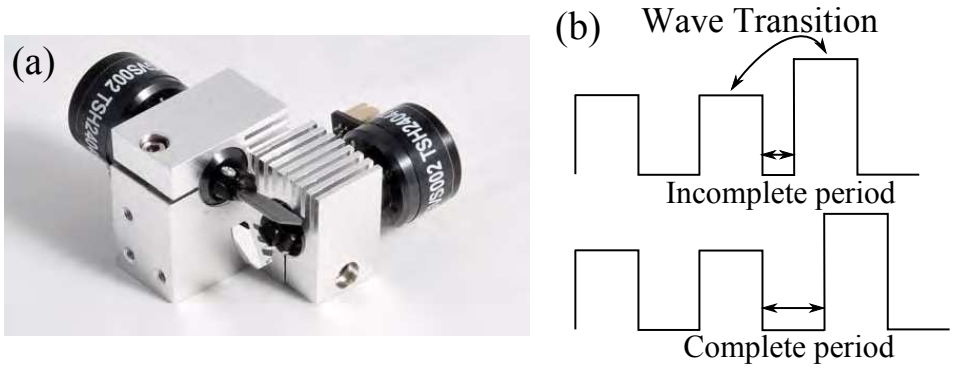


Figure 3.11: (a) GVS002 dual axis galvanometer scanning mirror system for time multiplexing multiple optical traps. Image courtesy of Thorlabs. (b)The smooth transition of the square wave amplitude is important to ensure proper synchronization of the multi-trap control signal and the events.

The galvanometer scanning mirror system is a high speed motor controlled mirror system capable of reflecting light to different directions, see Fig.3.11(a). The system consists of two main components: two motors that change the angles of the mirrors for the xy axes and detectors that feed back mirror positions to the system. The position of the laser spot can thus be adjusted on the fly by controlling the mirror position. The motor control requires analog voltage input thus a data acquisition card (National Instrument PCI6221) is disposed to generate the input signal.

As has been pointed out, the laser actuation mode is an important issue for force

feedback. The galvanometer is an optically bilateral system, where the illumination light goes through in exactly the opposite direction of the laser beam. DVS thus should be placed in front of the entrance of the laser into the galvanometer system, namely before the active actuator even comes into play, to have the laser spot constantly fixed in the scene, see Fig.3.8. In this manner, although laser is actively driven, laser center is fixed in DVS views and can be obtained in a calibration step.

Attention should be paid during the installation of the galvanometer cube system due to the effect of the change of the light polarization by galvanometer mirrors. The laser source used in our system is a linearly polarized one of 532nm wavelength and a polarized cube is used to separate the back focal plane light rays into different cameras. Therefore, respecting the correct wave polarisation is critical. If faulty, one way to correct it is to rotate the galvanometer cube and then redesign the affected optical path.

Light intensity is of great importance to the successful use of DVS. The optical path from the light source to DVS is lengthy in the current design. The light intensity is reduced to a significant extent by numerous optical components, such as the objective, the polarized cube, the galvanometer mirrors, the dicroic mirror and the laser filter located in front of DVS. Since DVS's bandwidth is strictly related to the light intensity, a dim light will not be able to guarantee a sufficient bandwidth. Therefore, a LED of very strong intensity (OSRAM,+2.17W) and of hot white color is chosen.

3.3.2 Events-Trap Correspondence

Two traps are generated by the galvanometer scanning mirror system. Time multiplexing method requires to synchronize the control signal of the galvanometer to the scene viewed by DVS. In other words, correspondences must be established between the switching moment of the mirror positions and the switching moment of DVS views of the scene.

Square waves of 250 Hz are used to continuously switch the mirror positions to create multiple traps. By changing the high and low electrical level of the square wave, one changes the position of the trap correspondingly. Time multiplexing means that time are equally shared between multiple traps. One incomplete wave period such as that shown in Fig.3.11 (b) will cause unbalanced time sharing thus ruin the whole synchronization process. This problem will actually occur if the transition of square wave amplitude to another value is not smooth. In the current setup, the wave generation command is conveyed by a non real-time operating system, so the clock task has to be completely attributed to hardware. The exact moment of amplitude transition hence cannot be fully controlled by software, which justifies the use of a hard real-time operating system in the near future. At the moment, this default has been partially corrected in a special synchronization procedure by the combined use of DVS's synchronization signal and the timestamps of events.

DVS is equipped with an external synchronization pin that has the ability to inject synthetic signals into the event stream on each falling edge detected on the

pin, which itself connects to the wave generation clock. The synchronization signal is thus generated at the exact moment of the trap swap. In the current DVS prototype, a synchronization signal is represented in the form of a status bit attached to the generated event data. Due to software implementation reasons, e.g. the use of socket protocol UDP that provides only best-effort delivery service, the successful recording of each falling edge is not guaranteed. To conclude, the DVS synchronization signal provides very useful hints to compensate the system timing default but cannot be used alone to establish the correct event-trap correspondence.

The switching frequency known a priori, the timestamps, clocked by a precise hardware counter in DVS, can be the most essential information to separate DVS's views from different traps. In other words, according to the scanning frequency, timestamp of each event is checked, which acts like a software clock running parallel and independently with the external trap switching clock. However, software clock suffers from screw and jitter of timings, the calculation of the precise switching moment will degrade due to the accumulation of timestamp errors in the long term. In other words, the software clock time will deviate more and more from the hardware wave generation clock during time. Hence, our synchronization strategy is the combined use of timestamps and the synchronization signals of DVS. The whole event-trap correspondence establishing procedure is illustrated in Algorithm 3.

Algorithm 3 Synchronizing DVS views with different laser traps

```
for every incoming event, do
    if there is a synchronization signal attached, then
        mark as the beginning of a switching period.
    if a period is finished according to timestamps, then
        switch the event-trap correspondence.
```

3.3.3 Multi-trap Experimental Results

In time-multiplexing force measurement, the establishment of events-trap correspondence is the most important issue. For the sake of the best detection, experiments have been conducted to carefully examine the details in event generation during trap switching. Several major differences from the single trap situation can be observed.

Firstly, it will be shown that only one type of event polarity will be retained for processing. The frame image of Fig.3.12 (c) shows a scenario where two beads are trapped simultaneously close to an obstacle. The two switching scenes enclosed in rectangles have their corresponding DVS views shown in Fig.3.12 (a) and (b), in which events are accumulated every 5 ms. As illustrated, ON and OFF polarities dominant in a "conjugate" moment of switching. More intuitively speaking, DVS view of scene 1 may possibly appear events of scene 2 with the opposite event polarity. Therefore, processing one type of event alone will largely eliminate the problem of undesirable including of the information due to polarity. However, this elimination is not a complete solution. A small quantity of traces will still be left

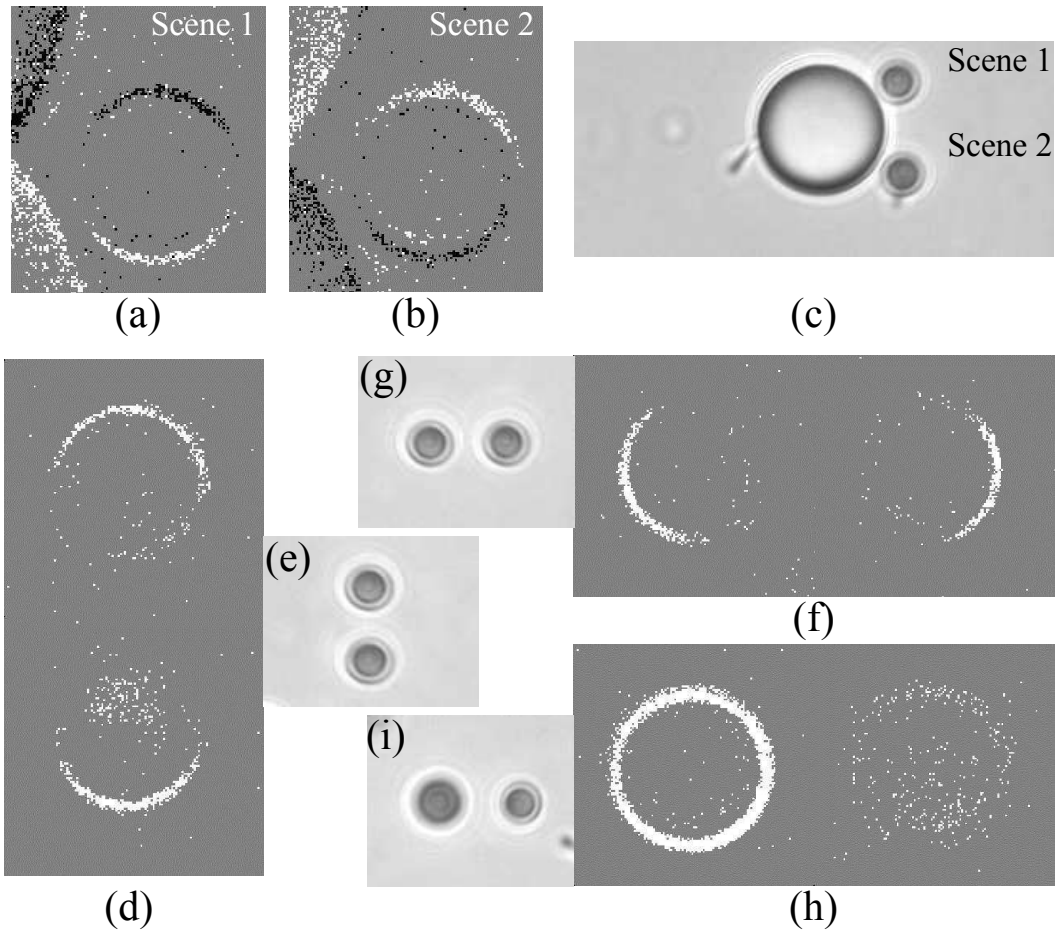


Figure 3.12: Images (a)∼(b) shows the DVS views during trap swap. Events accumulation time is 5 ms. ON and OFF polarity dominance occur in a conjugate moment of switching. (c) The enclosed areas in the classical frame image correspond to DVS views (a) and (b). (d)-(i) are three groups of DVS views and their corresponding frame images to show the impact of including light intensity difference in the event stream.

over. Experiments show that the higher the transition frequency of the galvonometer mirrors, the larger the quantity of undesirable traces be left in another scene. This adds a new type of noise to the current force measurement.

Unlike the situation of a single trap, the silicon retina DVS records this time not only the contour of the moving object but also the switching of the entire scene. In short, the DVS events generated by the new setup contain the following information:

- Movements of the beads;
- Intensity difference (scene A -scene B) and (scene B -scene A);

in which the last item is undesirable for bead position detection.

To illustrate what problems it can cause, imagine a moment when the Brownian motion is extremely weak, the events will then be generated purely due to scene difference. And if the two beads are located in exactly the same place, there will be no difference between the two scenes thus no event will be generated. Practically, beads do move and have differences. Fig.3.12(d), (f) and (h) are DVS views of two simultaneous trapped beads with OFF events eliminated. The undesirable generation of events by intensity difference between two scenes persists. Fig.3.12(e), (g) and (i) are frame images corresponding to the DVS views (d), (f) and (h) respectively. It can be seen that if two spheres are placed either vertically (case (d)(e)) or horizontally (case (f)(g)), the events will be more likely generated on the vertical side or the horizontal side of the contour, respectively. The dominant generation of events is in the same direction as that of the DVS view switching, where scene differences are more likely to appear. To further verify, an extra test is performed by trapping two spheres of different sizes, see (h) and (i). One observes that the events are mostly generated by the big sphere, which completely hides the smaller one during scene transition. As a result, how to strictly separate the movement of objects from the ego movement of the camera (scene background differences) becomes an interesting issue in temporal contrast change vision sensors as DVS.

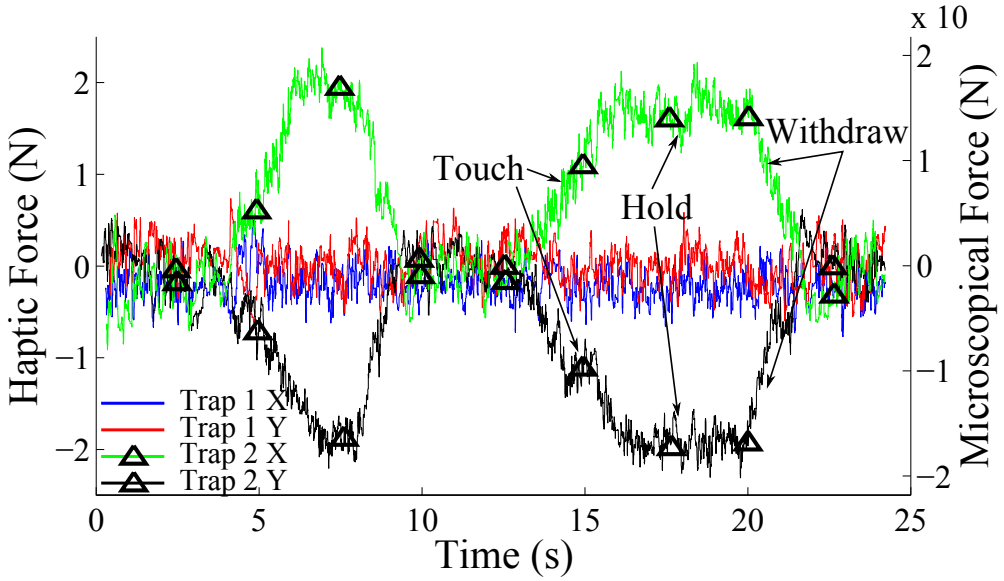


Figure 3.13: The forces (XY axes separated) of two beads in two optical traps when one bead is touching a target object, holding and being withdrew from the target, while another bead is left still. The inset image shows a photo of the scene.

Although several difficulties exist, our circle fitting algorithm is still robust enough to track circles of incomplete contour or contours that are not sharp. Further experiments are conducted to show the parallel force-feedback provided by tracking. Two beads are trapped simultaneously in two optical traps. Since only one haptic

device is available, it is used to "actively" control the position of one of the microsphere and left another one stay still (though it still exhibits Brownian motion!). Simultaneous control of both traps is possible simply by changing to a more adaptive haptic device. The experimental scene is shown in Fig.3.13 inset. The bead in Trap 1 is left unmoveable. The bead in Trap 2 comes into contact with a big sphere of diameter 10 μm several times. The generated forces are illustrated in Fig.3.13. One can see that the bead of Trap 1 is purely under Brownian motion. The second bead goes to touch the object, hold on for a moment and then withdraws from the target. This procedure repeats twice in the curves. This experiment clearly demonstrates that high speed force feedback on multiple-trap optical tweezers is completely plausible, although numerous aspects can still be improved, such as the inefficiency of optical setup, the imprecision of synchronization and the perturbations lying on the events processing.

3.4 Marketability

Optical tweezers and state-of-art measurement techniques provide new capacities for non-contact micromanipulation and micro-assembly. Haptic teleoperated optical tweezers, as this work demonstrated, show their strong potential of becoming a scientific exploration and manipulation tool. They provide perceptions of dynamic interactions with cells as if we were perceiving the stiffness of our skin with our fingers. The exploration of three-dimensional biological samples and the assembly of microscopic three-dimensional objects by the assistance of a haptic device are getting closer to a marketable reality.

Commercial optical tweezers are available, such as those in Elliot Scientific, Molecular Machines & Industries Inc. Their systems are able to control positions of multi-trap optical tweezers. However, no force measurement and haptic feedback are possible.

Technologies at Tholabs can read data from quadrant photo detector (QPD) and camera, but still "does not include any (software) routines that will analyse the data and calculate strength / stiffness values", as suggested in its manual. In other words, although it provides force sensor data for off-line processing, it does not address the problem of real-time force measurement.

Carl zeiss's "PALM MicroTweezers" allows the quantitative measurement of forces by purchasing an additional software module specialized at image processing. The limitations of frame based processing need not be emphasized again. JPK instrument has recently announced a new product "NanoTracker^{TM} 2" that provides high speed force measurement by the method of analysing the data of 3D QPDs. "NanoTracker^{TM} 2" allows precise force measurement but is difficult to be coupled with haptic devices fulfilling arbitrary tasks that involve object contact or overlapping. In other words, the commercial force modules are still not able to achieve speed and robustness double win.

This thesis provides the dedicated solution and demonstrates that the speed and

Figure 3.14: NanoTracker^{TM} 2 setup on Zeiss Axio Observer. Image courtesy of JPK Instrument.

robustness can be a double victory. The presence of haptic feedback significantly strengthens the usability and repeatability of the existing optical tweezers systems, to the extent of being a permanent demonstration in ISIR laboratory. The speed and the dexterity of the operator have been greatly improved thanks to the high bandwidth and high fidelity of the force feedback. A patent has been submitted for this system. The software including the tracking algorithm could be a supplement to the existing commercial optical tweezers systems to supply additional functions of force measurement and real-time haptic assisted manipulation in complex environment. This technology has strong potential to fulfil the gap of the current market vacancy.

Conclusions

In order to achieve dexterous 3D manipulation in a complex environment by optical tweezers, the system should be able to provide 3D force measurement. Several methods have been proposed to expand the 2D microsphere tracking capability to the third dimension, including averaging intensities on frame images, polarity integration, 3D Hough transform and robust circle fitting. The limitations and usability of each method have been studied and compared. The circle fitting algorithm is finally selected as the most suitable solution for our haptic feedback optical tweezers system.

A three-dimensional haptic feedback system suitable for multi-trap optical tweezers has thus been realized. This is the ultimate design to remove the current limi-

tations in the fields of micromanipulation, micro-assembly and microbiology exploration. A reproducible experiment of a complex three-dimensional exploration allows an unprecedented shape sensation of a spherical micro-object thanks to our tactile level quality of force feedback. Multi trapping is also achieved by a galvanometer using a time multiplexing method. DVS data is synchronized with the rotation of the mirrors of the galvanometer in order to achieve simultaneous multiple-sphere position detection. Finally, the marketability of the system is analysed to envision a profitable future.

Up to now, the first application - a complete high speed haptic feedback optical tweezers system - has been accomplished. In the next chapter, the second application - stable haptic virtual assistance of a microgripper based system - will be presented. The difficulty lies in that 1 kHz is required to ensure the stability of such system. Moreover, the form of the microtool is arbitrary and more complex than the microsphere. No analytic equation can be easily employed to describe the form. To meet the demand, an original event based shape registration algorithm sensible to the shape of the tool will be designed and integrated into a new micromanipulation system.

Conclusions and Perspectives

Silicon retina is a cutting-edge vision sensor that records highly compressed sparse data that retains only temporal contrast or moving edges. Event based vision focuses mainly on the topic of processing the data from silicon retina. The existing event based algorithms and successful applications are rare due to its novelty. This work is *one of the first* theses trying to apply the event based sensor and algorithms to serious robotic applications, and *the first* to microrobotics. More specifically, two haptic feedback teleoperated micromanipulation systems have been developed.

As the title of the thesis suggests, the first part of the contribution is a series of vision algorithms with the decoration term *the asynchronous event based*. They are to perform high speed circle tracking in two dimensional (Chapter 2), three dimensional (Chapter 3) space or shape registration (Chapter 4). Different from frame based visual paradigm, the asynchronous event vision sensor samples the (x, y, t) spatio-temporal space hence results in an efficient input for various algorithms. The sparse nature of visual data reduces the redundant information and retains only the dynamics. This allows for a lightweight computational load as has been shown. The presented event-based computation to some extent captures the asynchronous and time-continuous computation presented in biological neural systems. The computation is data driven thus provides a natural robustness to static occlusions. These advantages are the principle reasons that the event based vision is employed to micromanipulation applications, where high speed processing is mandatory. As a result, they allows circle tracking and shape registration at an equivalent frame-rate of much more than the required 1 kHz.

The second part of the contribution of the thesis comes to the microrobotic applications, which are the force feedback optical tweezers system and the haptic virtual assistance of microgripper system.

Chapter 2 and 3 construct one application in two steps. In the first step, we have focused on the two dimensional force feedback in order to ensure the full plausibility of DVS on the teleoperated optical tweezers system. A high speed circle detection method – the Continuous Hough Transform has been introduced which is shown to be more robust than the existing high speed microparticle tracking methods. High dynamic Brownian motion detection has been performed and a maximum information refreshing rate upto 30 kHz has been achieved. The developed system demonstrates for the first time the possibility of high bandwidth reliable piconewton level force sensation. The requested sampling frequency of 1 kHz has been successfully achieved. To validate the method, the trapped bead has been used as a tool to touch an irregular object and to explore a complex surface.

In the second step, the system is made full-fledged to the whole three dimensional space and to allow multi-trapping. After examining various 3D position tracking methods, the robust circle fitting algorithm has finally been selected as the most suitable one to perform monocular 3D tracking. A three-dimensional haptic feedback

system capable of generating forces in 3D and for multi-traps has been designed and then realized, all guarantee the high frequency requirement.

The goal of Chapter 4 is to provide a useful haptic feedback in the virtual assistance of a piezoelectric microgripper system. In order to calculate the forces on the complex form of a microgripper, we firstly deepen the theory of asynchronous high speed tracking by introducing a new event-based shape registration algorithm, which is able to track arbitrary form objects. A large quantity of experiments in indoor and outdoor environment has been conducted to validate the approach. To build the vision system of the microgripper system, the asynchronous tracking has been combined with a conventional frame based processing solely to track static objects. Experiments of pick-and-place of a glass sphere of 50 µm has been conducted. As the results shown, the frequency of the haptic feedback stays constantly higher than the required 1 kHz to guarantee stability and the impact of the sampling frequency on the stability of such system has been carefully analysed.

To resume, the low processing speed of conventional camera images and the zero tolerance of asymmetry by photo-diode constitute serious bottlenecks in reliable vision based haptic feedback teleoperation systems. In this thesis, new concepts of asynchronous event based vision have been established as a promising solution and successfully been brought to two concrete force feedback micromanipulation applications.

Concerning the perspectives, the most important future advancement will be the high bandwidth three dimensional force feedback on multi-trap optical tweezers. This thesis has shown that simultaneous force measurements on two trapped beads are plausible. However, the use of a galvanometer has two major disadvantages. Firstly, switching mirrors limit the motions in two dimensional space. Secondly, time-sharing strategy is not very energy efficient thus the number of beads that can be trapped is restrained, typically around ten. Spatial Light Modulator (SLM) is an advantageous way in creating multiple optical traps. It enables complex interactions between a computer and the intensity (amplitude modulation) and phase (phase modulation) characteristics of a light beam. SLM modulates the optical traps in three dimensional space and its phase modulation is very efficient in terms of energy. The fast hologram generation during phase modulation can be ensured by the state-of-the-art GPU parallel processing. Moreover, by changing the experimental setup, stereoscopic vision system can also be envisioned for enhanced position detection in the axial direction. Therefore, high speed 3D force measurement on multi-trap optical tweezers using SLM will be the next important advancement in the current micromanipulation technology.

The possible degree(s) of freedom in the manipulation process can be further extended. Rotation can be produced on either a non-spherical object or by an aberrant laser beam. High speed detection of the object rotation provides the mean to feedback also the momentum of the trapped object. In this case, the flexibility of the optical tweezers advances one big step further thanks to the extra degree(s) of freedom.

The immensely improved capabilities of optical tweezers offer new approaches for

micro/nanomanipulation. The high degrees of freedom multi-trap optical tweezers system with highly transparent force feedback may be beneficial in three different scenarios in the future. The first scenario is the ultra-flexible "microhand". Simultaneously controlled beads act as fingers of a human hand in the microworld. They can then be used to hold and manipulate complex biomedical samples.

The second scenario is the "microtool end-effector adaptor". In this scenario, the trapped beads are not meant to manipulate other object directly but are used to hold mechanical microtools of more complex structures. The micro-tools are the real end-effectors but they usually do not include force sensors. The optical tweezers thus play the role of both the end-effector holder and the infield force sensor. The forces exerted on the complex structured tool will be distributed on each attached bead. Therefore the force exerted on the micro-tool can be calculated by combining the forces on the trapped beads. Different forms of micro-tool end-effectors can be attached to different configurations of traps. This will surely open a brand new way for infield force sensing without using complex and expensive structures on microtool end-effectors.

The third scenario is the "position clamping". The current technology for position clamping is to use QPDs thus limits the application to simple environment conditions. High speed and robust position detection is an important factor to clamp beads steadily in the desired locations thus will enable automatic optical tweezers systems that accomplish real-world tasks.

Optical tweezers with intuitive haptic interactions have the potential to be the next standard tool in biology or nanotechnology laboratories. In order to bring this vision to reality, one must-do work is to search real-world applications. By collaborating with other laboratory or industrial partners, benefits of the intuitive force feedback on biological, biomedical and nanotechnological applications should be continuously studied and demonstrated to public to populate the technology.

Up to the present, the perspectives mainly focus on the performance at the slave side of the teleoperation system. Micromanipulation is naturally a dexterous task thus requires the bandwidth of the haptic interface to be as high as possible. The present device Omega.3 is designed for a force sensibility of human arms (thus low frequency), while haptic devices such as Phantograph can be used for a sensitivity of human fingers (thus high frequency). A better idea will be to associate a vibrator (the highest frequency) to Phantograph, hence improve significantly the bandwidth at the master side of the teleoperation system. Furthermore, the price of an Omega device is so high that the users are limited to laboratory researchers. Hence, building high transparency while low cost haptic devices is another interesting topic in future research.

Appendix A

Publication List

Published:

Z.Ni, C.Pacoret, R.Benosman and S.Régnier (2013). 2D High Speed Force Feedback Teleoperation of Optical Tweezers. IEEE International Conference on Robotics and Automation. Pages 1700 - 1705

Z.Ni, A.Bolopion, J.Agnus, R.Benosman and S.Régnier (2012). Asynchronous Event-based Visual Shape Tracking for Stable Haptic Feedback in Microrobotics. IEEE Transactions on Robotics (T-RO). Vol 28 No 5 Pages 1081 - 1089

Z.Ni, A.Bolopion, J.Agnus, R.Benosman and S.Régnier (2012). Stable Haptic Feedback based on a Dynamic Vision Sensor for Microrobotics. International Conference on Intelligent Robots and Systems. Pages 3203 - 3208

Z.Ni, C.Pacoret, R.Benosman, S.Ieng and S.Régnier (2011). Asynchronous Event Based High Speed Vision for Micro-particles Tracking. Journal of microscopy. Vol 245 No 3 Pages 236 - 244

Pending:

Z.Ni, C.Pacoret, R.Benosman, and S.Régnier (2013) First Reliable Tactile Three-dimensional Exploration of Microparticle Shapes Using Optical Tweezers (Journal)

Z.Ni, S.Ieng, C.Posch, S. Régnier and R.Benosman (2013) kHz Visual Tracking using Event-based Time-Encoded Imaging (Journal)

Bibliography

[Abd El Munim 2007] H.E. Abd El Munim and A.A. Farag. *Curve/Surface Representation and Evolution Using Vector Level Sets with Application to the Shape-Based Segmentation Problem*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 6, pages 945 –958, june 2007. (Not cited.)

[Agnus 2009] J. Agnus, D. Hériban, M. Gauthier and V. Pétrini. *Silicon end-effectors for microgripping tasks*. Precision Engineering, vol. 33, no. 4, pages 542–548, October 2009. (Not cited.)

[Ammi 2007] M. Ammi and A. Ferreira. *Robotic Assisted Micromanipulation System using Virtual Fixtures and Metaphors*. In IEEE International Conference on Robotics and Automation, pages 454–460, 2007. (Cited on pages 31 and 32.)

[Andersen 2009] K. N Andersen, D. H Petersen, K. Carlson, K. Molhave, O. Sardan, A. Horsewell, V. Eichhorn, S. Fatikow and P. Boggild. *Multimodal Electrothermal Silicon Microgrippers for Nanotube Manipulation*. IEEE Transactions on Nanotechnology, vol. 8, no. 1, pages 76–85, January 2009. (Cited on page 28.)

[Arai 2000] F. Arai, M. Ogawa and T. Fukuda. *Indirect manipulation and bilateral control of the microbe by the laser manipulated microtools*. In IEEE International Conference on Intelligent Robots and Systems, pages 665–670, 2000. (Cited on pages 47 and 48.)

[Ashkin 1987] A. Ashkin, J.M. Dziedzic and T. Yamane. *Optical trapping and manipulation of single cells using infrared laser beams*. Nature, vol. 330, no. 6150, pages 769–771, 1987. (Cited on page 29.)

[Belongie 2002] S. Belongie, J. Malik and J. Puzicha. *Shape matching and object recognition using shape contexts*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 4, pages 509 –522, apr 2002. (Not cited.)

[Belshaw 2009] M.S. Belshaw and M.A. Greenspan. *A high speed iterative closest point tracker on an FPGA platform*. In IEEE 12th International Conference on Computer Vision Workshops, pages 1449 –1456, 27 2009-oct. 4 2009. (Not cited.)

[Benosman 2011] R. Benosman, S.H. Ieng, C. Posch and P. Rogister. *Asynchronous Event-based Hebbian Epipolar Geometry*. IEEE Transactions on Neural Networks, 2011. (Cited on page 22.)

[Benosman 2012] R. Benosman, S.H. Ieng, C. Clercq, C. Bartolozzi and M. Srinivasan. *Asynchronous Frameless Event-based Optical Flow.* Neural Networks, 2012. (Cited on pages 22 and 23.)

[Besl 1992] P.J. Besl and H.D. McKay. *A method for registration of 3-D shapes.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 2, pages 239 –256, Feb 1992. (Not cited.)

[Beyeler 2009] F. Beyeler, S. Muntwyler and B. J Nelson. *A six-axis MEMS force–torque sensor with micro-Newton and nano-Newtonmeter resolution.* Journal of Microelectromechanical Systems, vol. 18, no. 2, pages 433–441, April 2009. (Cited on page 28.)

[Boahen 2000] K. A. Boahen. *Point-to-point connectivity between neuromorphic chips using address-events.* IEEE Trans. Circuits Syst. II, pages 416–434, 2000. (Cited on page 12.)

[Bolopion 2009] A. Bolopion, B. Cagneau and S. Régnier. *2D micro teleoperation with force feedback.* In The 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 3265–3270, 2009. (Cited on page 31.)

[Bolopion 2010] A. Bolopion, H. Xie, S. Haliyo and S. Régnier. *3D Haptic Handling of Microspheres.* In 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010), pages 6131 – 6136, 2010. (Not cited.)

[Bolopion 2012] A. Bolopion, Hui Xie, D.S. Haliyo and S. Regnier. *Haptic Teleoperation for 3-D Microassembly of Spherical Objects.* IEEE/ASME Transactions on Mechatronics, vol. 17, no. 1, pages 116 –127, feb. 2012. (Cited on pages 28, 31 and 32.)

[Bolopion 2012a] A. Bolopion, Z. Ni, J. Agnus, R. Benosman, and S. Régnier. *Stable Haptic Feedback based on a Dynamic Vision Sensor for Microrobotics.* IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 3203–3208, Feb. 2012. (Not cited.)

[Bookstein 1989] F.L. Bookstein. *Principal warps: thin-plate splines and the decomposition of deformations.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 6, pages 567 – 585, jun 1989. (Not cited.)

[Bowman 2010] R. Bowman, G. Gibson and M. Padgett. *Particle tracking stereomicroscopy in optical tweezers: Control of trap shape.* Opt. Express, vol. 18, no. 11, pages 11785–11790, May 2010. (Cited on page 58.)

[Cappelleri 2009] D.J. Cappelleri, G. Piazza and V. Kumar. *Two-dimensional, vision-based μN force sensor for microrobotics*. In IEEE International Conference on Robotics and Automation, pages 1016–1021, Kobe, May 2009. (Cited on page 28.)

[Chaillet 2010] N. Chaillet and S. Régnier. Microrobotics for micromanipulation. J. Wiley & Sons, 2010. (Cited on page 28.)

[Challis 1995] J. H. Challis. *A procedure for determining rigid body transformation parameters*. Journal of Biomechanics, vol. 28, no. 6, pages 733 – 737, 1995. (Not cited.)

[Chang 2006] C. Chang, P. Hsiao and Z. Huang. *Integrated Operation of Image Capturing and Processing in FPGA*. International Journal of Computer Science and Network Security, vol. 6, no. 1, pages 173 –180, 2006. (Cited on page 24.)

[Chawda 2011] V. Chawda and M. K. O'Malley. *Vision-Based Force Sensing for Nanomanipulation*. IEEE/ASME Transactions on Mechatronics, vol. 16, no. 6, pages 1177 – 1183, 2011. (Cited on page 28.)

[Chen 2011] D. Chen, D. Matolin, Bermak A. and C. Posch. *Pulse Modulation Imaging - Review and Performance Analysis*. In IEEE Transactions on Biomedical Circuits and Systems, volume 5, pages 64–82, 2011. (Cited on page 14.)

[Cheong 2010] F. Cheong, B. J. Krishnatreya and D. G. Grier. *Strategies for three-dimensional particle tracking with holographic video microscopy*. Opt. Express, vol. 18, no. 13, pages 13563–13573, Jun 2010. (Cited on page 58.)

[Colgate 1997] J. Colgate and G. Schenkel. *Passivity of a class of sampled-data systems: Application to haptic interfaces*. Journal of Robotic Systems, vol. 14, no. 1, pages 37–47, 1997. (Cited on page 29.)

[Conradt 2009] J. Conradt, M. Cook, R. Berner, P. Lichtsteiner, R J. Douglas and T Delbruck. *A pencil balancing robot using a pair of AER dynamic vision sensors*. In IEEE International Symposium on Circuits and Systems, page 781, 2009. (Cited on page 21.)

[Constandinou 2006] T.G. Constandinou and C. Toumazou. *A micropower centroiding vision processor*. IEEE Journal of Solid-State Circuits, vol. 41, no. 6, pages 1430 – 1443, june 2006. (Cited on pages 24 and 36.)

[Cremers 2006] D. Cremers, S. J. Osher and S. Soatto. *Kernel Density Estimation and Intrinsic Alignment for Shape Priors in Level Set Segmentation*. Int. J. Comput. Vision, vol. 69, no. 3, pages 335–351, September 2006. (Not cited.)

[Crocker 1996] J. C. Crocker and D. G. Grier. *Methods of Digital Video Microscopy for Colloidal Studies*. Journal of Colloid and Interface Science, vol. 179, no. 1, pages 298 – 310, 1996. (Cited on page 44.)

[Dahmouche 2010] R. Dahmouche, N. Andreff, Y. Mezouar and P. Martinet. *Efficient high-speed vision-based computed torque control of the orthoglide parallel robot*. In IEEE International Conference on Robotics and Automation (ICRA), pages 644 –649, may 2010. (Cited on page 24.)

[Delbruck 2007] T. Delbruck and P. Lichtsteiner. *Fast sensory motor control based on event-based hybrid neuromorphic-procedural system*. In IEEE International Symposium on Circuits and Systems, pages 845 – 848, 2007. (Cited on pages 20, 44 and 46.)

[Delbruck 2008] T. Delbruck. *Frame-free dynamic digital vision*. In Proceedings of Intl. Symp. on Secure-Life Electronics, Advanced Electronics for Quality Life and Society, pages 21–26, 2008. (Cited on page 20.)

[Delbruck 2010] T. Delbruck, B. Linares-Barranco, E. Culurciello and C. Posch. *Activity-driven, event-based vision sensors*. In IEEE International Symposium on Circuits and Systems, pages 2426–2429, 2010. (Cited on page 12.)

[Duc 2008] Trinh Chu Duc, Gih-Keong Lau, J. F Creemer and P. M Sarro. *Electrothermal Microgripper With Large Jaw Displacement and Integrated Force Sensors*. Journal of Microelectromechanical Systems, vol. 17, no. 6, pages 1546–1555, December 2008. (Cited on page 28.)

[Fahlbusch 1998] S. Fahlbusch and S. Fatikow. *Force sensing in microrobotic systems-an overview*. In International Conference on Electronics, Circuits and Systems, volume 3, pages 259–262 vol.3, 1998. (Cited on page 28.)

[Fas] *http://advance.uri.edu/pacer/september2000/story9.htm*. (Cited on page 6.)

[Ferreira 2006] A. Ferreira and C. Mavroidis. *Virtual reality and haptics for nanorobotics*. Robotics Automation Magazine, IEEE, vol. 13, no. 3, pages 78–92, 2006. (Cited on page 31.)

[Fish 2007] A. Fish, L. Sudakov-Boreysha and O. Yadid-Pecht. *Low-power tracking image sensor based on biological models of attention*. Int. J. Information Theories and Applications, vol. 14, pages 103–114, 2007. (Cited on page 36.)

[Fornefett 1999] M. Fornefett, K. Rohr and H. Stiehl. *Elastic Registration of Medical Images using Radial Basis Functions with Compact Support*. In IEEE Conf. on Computer Vision and Pattern Recognition, volume 1, pages 1402–1409, 1999. (Not cited.)

[Fu 2008] Zhengming Fu, T. Delbruck, P. Lichtsteiner and E. Culurciello. *An Address-Event Fall Detector for Assisted Living Applications*. IEEE Transactions on Biomedical Circuits and Systems, vol. 2, no. 2, pages 88 –96, june 2008. (Cited on page 21.)

[Gibson 2008] G. Gibson, J. Leach, S. Keen, A.J. Wright and M. J. Padgett. *Measuring the accuracy of particle position and force in optical tweezers using high-speed videomicroscopy*. Optics Express, vol. 16, pages 14561–14570, 2008. (Cited on page 36.)

[Gibson 2012] G. M. Gibson, R. W. Bowman, A. Linnenberger, M. Dienerowitz, D. B. Phillips, D. M. Carberry, M. J. Miles and M. J. Padgett. *A compact holographic optical tweezers instrument*. Rev. Sci. Instrum., vol. 83, page 113107, 2012. (Cited on page 72.)

[Gosse 2002] C. Gosse and V. Croquette. *Magnetic Tweezers: Micromanipulation and Force Measurement at the Molecular Level*. Biophysical Journal, vol. 82, pages 3314–3329, 2002. (Cited on pages 37, 59 and 60.)

[Greminger 2004] M.A. Greminger and B.J. Nelson. *Vision-based force measurement*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 3, pages 290–298, March 2004. (Cited on page 29.)

[Grier 2003] D.G. Grier. *A revolution in optical manipulation*. Nature, vol. 424, no. 6950, pages 810–816, 2003. (Cited on page 29.)

[Haliyo 2003] D.S. Haliyo, S. Régnier and J.C. Guinot. *[mü] MAD, the adhesion based dynamic micro-manipulator*. European Journal of Mechanics-A/Solids, vol. 22, no. 6, pages 903–916, 2003. (Cited on pages 27 and 29.)

[Hartley 2003] R. Hartley and A. Zisserman. Multiple view geometry in computer vision. Cambridge University Press, 2003. (Not cited.)

[Hersch 2012] M. Hersch, A. Billard and S. Bergmann. *Iterative Estimation of Rigid-Body Transformations*. Journal of Mathematical Imaging and Vision, pages 1–9, 2012. (Not cited.)

[Hofstatter 2011] M. Hofstatter, M. Litzenberger, D. Matolin and C. Posch. *Hardware-accelerated address-event processing for high-speed visual object recognition*. In 18th IEEE International Conference on Electronics, Circuits and Systems (ICECS), pages 89 –92, dec. 2011. (Not cited.)

[Hough 1959] P. V. C. Hough. *Machine Analysis of Bubble Chamber Pictures*. In Proc. Int. Conf. High Energy Accelerators and Instrumentation, pages 554–556, 1959. (Cited on pages 15 and 37.)

[Huang 2006] X. Huang, N. Paragios and D. N. Metaxas. *Shape Registration in Implicit Spaces Using Information Theory and Free Form Deformations*. IEEE Trans. Pattern Analysis and Machine Intelligence (TPAMI, vol. 28, pages 1303–1318, 2006. (Not cited.)

[Huber 2009] P. J. Huber. Robust statistics. John Wiley & Sons Inc., 2009. (Cited on page 66.)

[Ioannou 1999] D. Ioannou, W. Huda and A. F. Laine. *Circle recognition through a 2D Hough Transform and radius histogramming*. Image and Vision Computing, vol. 17, pages 15–26, 1999. (Cited on pages 38 and 63.)

[Ivan 2011] I.A. Ivan, Gilgueng Hwang, J. Agnus, M. Rakotondrabe, N. Chaillet and S. Regnier. *First experiments on MagPieR: A planar wireless magnetic and piezoelectric microrobot*. In IEEE International Conference on Robotics and Automation (ICRA), pages 102 –108, may 2011. (Cited on pages 26, 27 and 46.)

[Jost 2003] T. Jost and H. Hugli. *A multi-resolution ICP with heuristic closest point search for fast and robust 3D registration of range images*. In Proceedings of Fourth International Conference on 3-D Digital Imaging and Modeling, pages 427 – 433, oct. 2003. (Not cited.)

[Kandel 1991] E. Kandel, J. Schwartz and T.M. Jessel. Principles of neural science. Elsevier, New York, 1991. (Cited on page 20.)

[Kim 2008] K. Kim, X. Liu, Y Zhang and Y. Sun. *Micronewton force-controlled manipulation of biomaterials using a monolithic MEMS microgripper with two-axis force feedback*. In IEEE International Conference on Robotics and Automation, pages 3100–3105, May 2008. (Cited on page 28.)

[Kim 2010] Daehwan Kim and Daijin Kim. *A Fast ICP Algorithm for 3-D Human Body Motion Tracking*. Signal Processing Letters, IEEE, vol. 17, no. 4, pages 402 –405, april 2010. (Not cited.)

[Lange 1998] F. Lange, P. Wunsch and G. Hirzinger. *Predictive Vision Based Control of High Speed Industrial Robot Paths*. In Proc. IEEE Int. Conference on Robotics and Automation, pages 2646–2651, 1998. (Cited on page 23.)

[Lansdorp 2013] B. M. Lansdorp, S. J. Tabrizi, A. Dittmore and O. A. Saleh. *A high-speed magnetic tweezer beyond 10,000 frames per second*. review of scientific instrument, vol. 84, page 044301, 2013. (Cited on page 36.)

[Lee 2012] J. Lee, T. Delbruck, P.K.J. Park, M. Pfeiffer, C.W. Shin, H. Ryu and B.C. Kang. *Live demonstration: Gesture-Based remote control using stereo pair of dynamic vision sensors,*. In ISCAS, pages 736–740, Seoul, 2012. (Cited on pages 18 and 21.)

[Lewalle 2007] A Lewalle. *How do muscles produce work? Using optical tweezers to study molecular machines*. science in school, vol. 4, pages 18–21, 2007. (Cited on pages 29 and 30.)

[Lichtsteiner 2008] P. Lichtsteiner, C. Posch and T. Delbruck. *A 128*128 120 dB 15us Latency Asynchronous Temporal Contrast Vision Sensor*. IEEE Journal of Solid-State Circuits, vol. 43, no. 2, pages 566–576, 2008. (Cited on pages 12, 13 and 16.)

[Litzenberger 2006a] M. Litzenberger, A.N. Belbachir, N. Donath, G. Gritsch, H. Garn, B. Kohn, C. Posch and S. Schraml. *Estimation of Vehicle Speed Based on Asynchronous Data from a Silicon Retina Optical Sensor*. In Intelligent Transportation Systems Conference. ITSC '06. IEEE, pages 653 –658, sept. 2006. (Cited on page 20.)

[Litzenberger 2006b] M. Litzenberger, C. Posch, D. Bauer, A.N. Belbachir, P. Schon, B. Kohn and H. Garn. *Embedded Vision System for Real-Time Object Tracking using an Asynchronous Transient Vision Sensor*. In Digital Signal Processing Workshop, 12th - Signal Processing Education Workshop, 4th, pages 173 –178, sept. 2006. (Cited on page 20.)

[Lopez-Walle 2008] B. Lopez-Walle, M. Gauthier and N. Chaillet. *Principle of a Submerged Freeze Gripper for Microassembly*. IEEE Transactions on Robotics, vol. 24, no. 4, pages 897–902, 2008. (Cited on page 28.)

[Mahowald 1992] M. Mahowald. *VLSI analogs of neuronal visual processing: a synthesis of form and function*. PhD thesis, California Institute of Technology, Pasadena, CA, 1992. (Cited on page 12.)

[Markovsky 2009] I. Markovsky and S. Mahmoodi. *Least-Squares Contour Alignment*. Signal Processing Letters, IEEE, vol. 16, no. 1, pages 41 – 44, jan. 2009. (Not cited.)

[Mead 1989] C.A. Mead. *Adaptive retina*. In Analog VLSI implementation of neural systems, 1989. (Cited on page 10.)

[Namiki 2003] A. Namiki, Y. Imai, M. Ishikawa and M. Kaneko. *Development of a high-speed multifingered hand system and its application to catching*. In Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on, volume 3, pages 2666 – 2671 vol.3, oct. 2003. (Cited on pages 23 and 24.)

[Neuman 2004] K.C. Neuman and S. M. Block. *Optical trapping*. Review of Scientific Instruments, vol. 75, no. 9, pages 2787–2809, 2004. (Cited on pages 29, 30 and 48.)

[Ni 2011] Z. Ni, C. Pacoret, R. Benosman, S. Ieng, and S. Régnier. *Asynchronous Event Based High Speed Vision for Micro-particles Tracking*. Journal of microscopy, vol. 245, no. 3, pages 236–244, 2011. (Not cited.)

[Ni 2012] Z. Ni, A. Bolopion, J. Agnus, R. Benosman, and S. Régnier. *Asynchronous Event-based Visual Shape Tracking for Stable Haptic Feedback in Microrobotics*. IEEE Transactions on Robotics, vol. 28, no. 5, pages 1081–1089, 2012. (Not cited.)

[Ni 2013] Z. Ni, C. Pacoret, R. Benosman, and S. Régnier. *2D High Speed Force Feedback Teleoperation of Optical Tweezers*. IEEE International Conference on Robotics and Automation, pages 1700–1705, 2013. (Not cited.)

[Noordam 2000] J.C. Noordam, G. W. Otten, A.J.M. Timmermans and B.H. van Zwol. *High Speed Potato Grading and Quality Inspection Based on a Color Vision System*, 2000. (Cited on pages 23 and 24.)

[Oku 2005] H. Oku, N. Ogawa, M. Ishikawa and K. Hashimoto. *Two-dimensional tracking of a motile micro-organism allowing high-resolution observation with various imaging techniques*. Review of Scientific Instruments, vol. 245, no. 3, pages 236–244, mar 2005. (Cited on page 10.)

[Onda 2012a] K. Onda and F. Arai. *Multi-beam bilateral teleoperation of holographic optical tweezers*. Optics Express, vol. 20, pages 3633–3641, 2012. (Cited on pages 48 and 71.)

[Onda 2012b] K. Onda and F. Arai. *Parallel teleoperation of holographic optical tweezers using multi-touch user interface*. In 2012 IEEE International Conference on Robotics and Automation (ICRA), pages 1069 –1074, may 2012. (Cited on pages 48 and 72.)

[Pacoret 2009] C. Pacoret, R. Bowman, G. Gibson, S. Haliyo, D. Carberry, A. Bergander, S. Régnier and M. Padgett. *Touching the microworld with force-feedback optical tweezers.* Optics Express, vol. 17, pages 10259–10264, 2009. (Cited on pages 28, 29 and 48.)

[Paragios 2003] N. Paragios, M. Rousson and V. Ramesh. *Non-rigid registration using distance functions.* Computer Vision and Image Understanding, vol. 89, no. 2–3, pages 142 – 165, 2003. (Not cited.)

[Posch 2008] C. Posch, D. Matolin and R. Wohlgenannt. *An asynchronous time-based image sensor.* In ISCAS, pages 2130–2133, 2008. (Cited on page 13.)

[Posch 2010] C. Posch. *High-DR Frame-Free PWM Imaging with asynchronous AER Intensity Encoding and Focal-Plane Temporal Redundancy Suppression.* ISCAS, 2010. (Cited on page 14.)

[Posch 2011] C. Posch, D. Matolin and R. Wohlgenannt. *A QVGA 143 dB Dynamic Range Frame-Free PWM Image Sensor With Lossless Pixel-Level Video Compression and Time-Domain CDS.* IEEE Journal of Solid-State Circuits, vol. 46, no. 1, pages 259 –275, jan. 2011. (Cited on page 13.)

[Reddy 2010] A. N. Reddy, N. Maheshwari, D. K. Sahu and G. K. Ananthasuresh. *Miniature Compliant Grippers With Vision-Based Force Sensing.* IEEE Transactions on Robotics, vol. 26, no. 5, pages 867–877, October 2010. (Cited on page 28.)

[Rogister 2011] P. Rogister, R. Benosman, S.H. Ieng, P. Lichtsteiner and T. Delbruck. *Asynchronous Event-based Binocular Stereo Matching.* IEEE Transactions on Neural Networks, 2011. (Cited on page 22.)

[Rueckert 1999] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach and D. Hawkes. *Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images.* In IEEE Transactions on Medical Imaging,, volume 8, pages 712–721, 1999. (Not cited.)

[Ruh 2011] D. Ruh, B. Tränkle and A. Rohrbach. *Fast parallel interferometric 3D tracking of numerous optically trapped particles and their hydrodynamic interaction.* Opt. Express, vol. 19, no. 22, pages 21627–21642, Oct 2011. (Cited on page 72.)

[Rusinkiewicz 2001] S. Rusinkiewicz and M. Levoy. *Efficient Variants of the ICP Algorithm.* In Third International Conference on 3D Digital Imaging and Modeling (3DIM), June 2001. (Not cited.)

[Saunter 2005] C. D. Saunter, G. D. Love, M. Johns and J. HOLMES. *FPGA technology for high speed, low cost adaptive optics*. the International Society for Optical Engineering, 2005. (Cited on page 36.)

[Serrano-Gotarredona 2009] R. Serrano-Gotarredona, M. Oster, P. Lichtsteiner, A. Linares-Barranco, R. Paz-Vicente, F. Gomez-Rodriguez, L. Camunas-Mesa, R. Berner, M. Rivas-Perez, T. Delbruck, Shih-Chii Liu, R. Douglas, P. Hafliger, G. Jimenez-Moreno, A.C. Ballcels, T. Serrano-Gotarredona, A.J. Acosta-Jimenez and B. Linares-Barranco. *CAVIAR: A 45k Neuron, 5M Synapse, 12G Connects/s AER Hardware Sensory System for High-Speed Visual Object Recognition and Tracking*. Neural Networks, IEEE Transactions on, vol. 20, no. 9, pages 1417 –1438, sept. 2009. (Not cited.)

[S.Keen 2007] S.Keen, J.Leach and G.Gibson. *Comparison of a high-speed camera and a quandrant detector for measuring displacements in optical tweezers*. J.Opt. A:Pure Appl.Opt., pages S264–S266, sep. 2007. (Cited on page 36.)

[Tamadazte 2010] B. Tamadazte, N. L. Piat and S. Dembele. *Robotic Micromanipulation and Microassembly Using Monoview and Multiscale Visual Servoing*. IEEE/ASME Transactions on Mechatronics, vol. PP, no. 99, pages 1 – 11, 2010. (Cited on page 27.)

[Uchikura 2006] R. Uchikura, T. Ishihara, T. Tsujita, T. Sugimura, J. Deguchi, M. Koyanagi and M. Uchiyama. *Development of a High Speed Vision System for Mobile Robots*. In IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 1372 –1377, oct. 2006. (Cited on pages 23 and 24.)

[Vis] *http://www.visionresearch.com/*. (Cited on page 6.)

[Watanabe 2007] Y. Watanabe, T. Komuro and M. Ishikawa. *955-fps Real-time Shape Measurement of a Moving/Deforming Object using High-speed Vision for Numerous-point Analysis*. In IEEE International Conference on Robotics and Automation, pages 3192 –3197, april 2007. (Cited on page 23.)

[Xie 2009] H. Xie and S. Régnier. *Three-dimensional automated micromanipulation using a nanotip gripper with multi-feedback*. Journal of Micromechanics and Microengineering, vol. 19, no. 7, page 075009, 2009. (Cited on pages 27 and 31.)

[Xie 2011] H Xie, C. D Onal, S Régnier and M. Sitti. Atomic force microscopy based nanorobotics, volume 71 of *Springer Tracts in Advanced*

Robotics. Modelling, simulation, setup building and experiments series édition, 2011. (Cited on page 28.)

[Yang 2011] J. Yang. *The thin plate spline robust point matching (TPS-RPM) algorithm: A revisit*. Pattern Recognition Letters, vol. 32, no. 7, pages 910 – 918, 2011. (Not cited.)

[Zhang 1997] Z. Zhang. *Parameter Estimation Techniques: A Tutorial with Application to Conic Fitting*. Image and Vision Computing Journal, vol. 15, no. 1, pages 59–76, 1997. (Cited on page 66.)

[Zhang 2008] Z. Zhang and C. Menq. *Three-dimensional particle tracking with subnanometer resolution using off-focus images*. Appl. Opt., vol. 47, no. 13, pages 2361–2370, May 2008. (Cited on page 58.)

[Zhang 2009] Y. L. Zhang, M. L. Han, J. Vidyalakshmi, C. Y. Shee and W. T. Ang. *Automatic control of mechanical forces acting on cell biomembranes using a vision-guided microrobotic system in Computer Microscopy*. Journal of Microscopy, vol. 236, no. 1, pages 70–78, 2009. (Cited on page 28.)

[Zhang 2010] Y. Zhang, B.K. Chen, X.Y. Liu and Y. Sun. *Autonomous robotic pick-and-place of micro objects*. IEEE. Trans. Robotics, vol. 26, no. 1, pages 200–207, 2010. (Cited on page 28.)

[Zheng 2006] Y. Zheng and D. Doermann. *Robust point matching for nonrigid shapes by preserving local neighborhood structures*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, page 2006, 2006. (Not cited.)

[Zitovà 2003] B. Zitovà and J. Flusser. *Image registration methods: a survey*. Image and Vision Computing, vol. 21, pages 977–1000, 2003. (Not cited.)