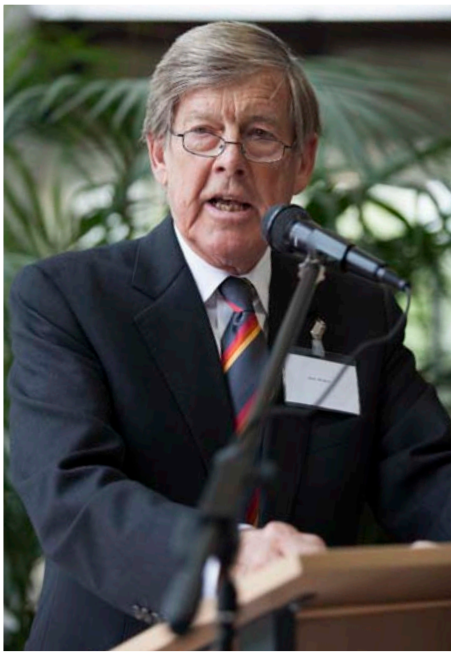
The Ring

The journal of the Cambridge Computer Lab Ring

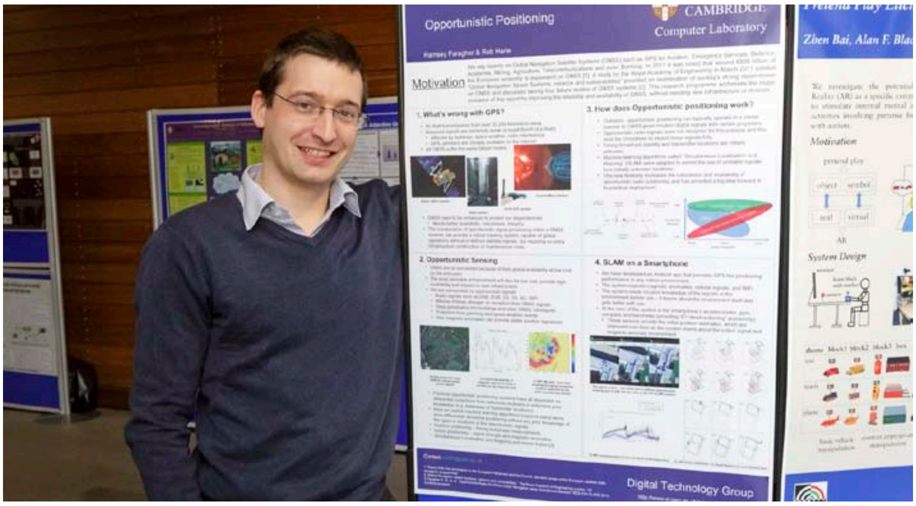
Issue XXXIII — May 2013

The Computer Laboratory

Celebrating the first 75 years on 24th April 2013



Anticlockwise (from top left): Professor Sir Tony Hoare delivers the Wheeler Lecture; Professor Andy Hopper presents Dr Mike Lynch with a commemorative certificate; Professor Ahmed signs a copy of Cambridge Computing: The First 75 Years *for Professor Bjarne Stroustrup; Lord Broers addresses over 300 graduates and friends of the Computer Laboratory; Professor Haroon Ahmed, Professor Andy Hopper and Lord Broers at the book launch.*



Anticlockwise (from top left): Innovation Discussion Panel shares a joke with the audience (left to right: Dr Mike Lynch, Dr Mike Muller, Professor Andy Hopper, Dr Andy Harter, Dr Eben Upton); Poster competition winner Ramsey Faragher for his poster "Opportunistic positioning"; raising a glass to the Computer Laboratory at the Cambridge Ring annual dinner; Lord Broers delivers guest speech (left to right: Sir Robin Saxby, Dr Gerard Bricogne, Dr Hermann Hauser CBE, Professor Andy Hopper CBE); Professor Andy Hopper presents Lord Broers with a copy of Cambridge Computing: The First 75 Years.

Hall of Fame Awards

Address by Master of Ceremonies, **Professor Andy Hopper**

Today we have been celebrating the Computer Laboratory's 75th anniversary. It is a celebration not only of the Lab's academic excellence but also its role as a pre-eminent source of entrepreneurial energy.

The Hall of Fame Awards acknowledge both strands so, with this in mind, I'd like to start with the award for Publication of the Year and invite Professor Robinson to accept the award for "3D Constrained local model for rigid and non-rigid facial tracking" by Tadas Baltrušaitis, Peter Robinson and Louis-Philippe Morency. *[Read the paper on p13]*

Professor Peter Robinson receives the award from Lord Broers

Moving on to the categories which recognise those companies founded by Lab graduates and staff, the nominations for Product of the Year are:

Bromium for Bromium vSentry, Cronto for CrontoSign, RealVNC for VNC Automotive and Swiftkey for SwfitKey3.

Like last year, there was a very tight contest for the top spot and, like last year, this year's runner-up is SwiftKey. SwiftKey 3 was the world's best-selling Android app of 2012; it was downloaded over five million times and has more than two million active users. Over the past year it has won many accolades including "Most Innovative Mobile App" at the GSMA awards, a People's Voice award for Innovation at the Webbys, and the Recombu's "Best App 2012".

Pipping SwiftKey on the line for Product of the Year 2013 is a product that allows drivers and passengers to automatically detect, access and control virtually any mobile device or desktop computer from a vehicle's touch-screen, fixed input device (such as steering wheel switches) or by voice command.

Andy Harter accepts the award for Product of the Year

So, I'd like to invite Andy Harter, co-founder of RealVNC, to accept the award for RealVNC's VNC Automotive.

And now for Company of the Year 2013.

Over 200 companies have been started by Lab graduates and staff. This year, the judges' task of finding a winner from such a high-quality list of entries was no mean feat.

The Ring, Issue XXXIII, May 2013
Web: http://www.cl.cam.ac.uk/ring/
E-mail: cam-ring@cl.cam.ac.uk
Tel: +44 1223 763585
Post: William Gates Building, Cambridge CB3 0FD

Published three times a year. Copy deadline for the September 2013 issue is August 1st 2013.
The Ring is the journal of the Computer Lab Ring, which is the graduate association of the University of Cambridge Computer Laboratory.

Eben Upton (left) accepts the award on behalf of Raspberry Pi.

The nominations for Company of the Year are:

Bango, Cronto, Fusepump, Raspberry Pi, Swiftkey, Ubisense, Xsilon.

Bango has become a pioneer of mobile Internet payments. 2012 was the year that propelled the company into the big time. At the start of the year, Bango captured headlines with announcements of partnerships with Amazon and then Facebook. Next came an agreement with Microsoft to support its roll-out of Windows Marketplace. In the final quarter of 2012, Bango announced that its first integration with Google Play had gone live.

Cronto addresses the growing problem of fraud in Internet payments.

In 2012, Cronto's innovative visual transaction signing solution, that enables financial institutions to counter sophisticated Trojan and "Man in the Browser" attacks, was deployed by two leading European banks, Raiffeisen in Switzerland and Commerzbank in Germany.

In 2009, Robert Durkin and Chris Conn got a bank loan and started Fusepump to provide e-commerce marketing solutions to on-line retailers. Four years on the company employs 55 people in the UK, has clients across the world and has grown at more than 70% year-on-year.

Fusepump has had some notable successes: its technology has helped Nokia to generate €1 billion from sales of handsets, John Lewis to improve its on-line search revenues by 400% and achieve bumper sales for the past two Christmases, and Tesco and Sky to develop affiliate marketing across their brands. It is currently working with ASOS, the global on-line fashion and beauty retailer, to develop a closed affiliate marketing platform and advertising tools.

Raspberry Pi has recently celebrated its first birthday. This basic computer, which costs around £25, has taken the world by storm. It was designed primarily to get children interested in programming and inspire a new generation of British innovators. More than a million have been sold since orders started being accepted on February 29th 2012 — an amazing achievement for a computer made by a not-for-profit organisation.

2012 was a big year for SwiftKey. Their team grew from 30 to over 80, they closed deals with more than ten mobile manufacturers, including four of the top ten. They signed deals with 15 US healthcare providers for "SwiftKey Healthcare", shipped embedded software on some of the year's most popular smartphones and won "Startup of the Year" at the Guardian digital innovation awards.

Ubisense celebrated a number of significant events in 2012. The year marked its tenth anniversary, during which time the company has grown from being a handful of academic researchers to almost 200 people based on three different continents — though their HQ and R&D facilities are still firmly based in Cambridge. 2012 saw the anniversary of its flotation on AIM. Unlike other more high profile flotations, Ubisense has never dropped below its original IPO price. And the company achieved the distinction of winning not one but two Queen's Awards — recognising the company's on-going innovation and success in international trade.

Xsilon, a fabless semiconductor company, epitomises the idea of a lean startup. Despite receiving only £260k in funding its team now numbers eight, and it has built a technology demonstrator of its groundbreaking comms technology. Xsilon showcased its Hanadu in-home M2M solutions at UK Trade and Investment's British Business Embassy during the London 2012 Olympics and Paralympic Games.

So, as you can see, the judges were presented with an incredibly strong and varied list of entries and, despite the challenging task it presented, they have enjoyed the experience.

Sadly, there can only be one winner. By way of introduction, I'd like to read a quote from one of the co-founders:

"The younger generation has demonstrated significant intrigue in learning how to build and program their own computer devices. I have seen projects from Twittering chickens to home beer brewing kits being created using the Raspberry Pi and its accessories."

And with that, I'd like to invite Eben Upton, co–founder of the Raspberry Pi Foundation, to receive the award for Company of the Year 2013.

The end of an era

Margaret Levitt prepares to bid farewell to the Computer Laboratory

"The Faculty of Computer Science & Technology is seeking to appoint a Secretary of the Department in the Computer Laboratory following the forthcoming retirement of the current role holder." So read the notice that left all in no doubt that Margaret Levitt, who has acted as the Department's glue for the last 26 years, was hanging up her gown.

The Secretary of the Department is one of the three senior members of staff supporting the Head of Department; the other two being the Deputy heads. The role today is broad, covering department administration, personnel and building management — very different from the job offered in 1987.

It is largely thanks to Margaret's husband that she joined the Computer Laboratory. Indeed when offered the post of Student Administration Secretary, she turned it down.

Margaret was working at St Andrew's Junior School in Cambridge, as School Secretary, when she sent in her job application. She produced it on the school's only BBC Micro — rather impressive when most were still using typewriters to produce their CVs. Invited for interview, she arrived smartly dressed in silk shirt and attendant attire and presented herself to an interview panel dressed in shorts and sandals! Not only was the panel's garb somewhat startling for someone used to children in uniform and well-turned-out staff but, when the late Judy Bailey lifted up Margaret's CV by the corner and, adopting an intimidatory tone, said "What is this?", Margaret was taken aback. She left the interview, having been asked nothing about her experience; the interview consisted of talk about music.

Margaret was offered the job later that day but turned it down. How could she possibly work at such a "mad place"? Mr Levitt's attitude was less equivocal; "You could do the job standing on your head". So, after a night's reflection, Margaret phoned Judy Bailey to say she would accept the job if it was still available.

When Margaret joined the Computer Laboratory in 1987 most of the department's administrative needs were met by the Computing Service so Margaret's responsibilities were focused on student administration (predominantly diploma students as the full three-year tripos didn't begin until 1989). In 1990, Margaret felt she needed something more challenging and applied for the post of Faculty Secretary at Classics. Jean Bacon, who joined the Laboratory as its first female lecturer in 1985 and was acting Head of Department while Roger Needham was on one of his regular consulting visits to Digital Equipment Corporation's Systems Research Center in Palo Alto, got wind of Margaret's application and urged Roger to take defensive action to prevent Margaret from leaving.

Roger was easily persuaded and submitted a case to the Council for a Departmental Secretary. While the establishment of the post was making its passage through the University, Margaret was appointed, in May 1990, as Administration Officer 'in an unestablished capacity to undertake administrative duties in the Computer Laboratory'.

Margaret didn't spend long in this interim role as she was employed as Secretary of the Computer Laboratory from March 1st 1991. As well as being responsible for departmental administration including undergraduate teaching and graduate students, Margaret was appointed Secretary of the Computer Science Syndicate (departmental and teaching administration) and Secretary of the Degree Committee (dealing with graduate students).

It was fortunate that Margaret approached new challenges with gusto as she would soon be heavily involved in the planning for the new building.

In 1996, Roger Needham resigned the headship of the Computer Laboratory and took up the new office of Pro Vice-Chancellor. He was soon approached by the Microsoft Corporation with the proposal that he should establish a research laboratory for them in Cambridge — Microsoft's first non-US research lab. As part of the deal, the University received a US$20 million donation from the Gates Foundation for a new computer laboratory. There was much excitement that the Computer Laboratory, then housed in cramped conditions at the New Museums Site, would be moving into specially designed premises.

Under the headship of Robin Milner, Margaret was now fully engaged with planning. As anyone who has had building work done knows, it is a fraught time; Margaret was busy from the time the Building Group convened in June 1997 until the new building's topping out ceremony on October 31st 2000 when Ian Leslie, who had taken over as Head of Department in 1999, tightened the last bolt on a roof beam.

The creation of the new building marked the point at which the Computer Laboratory split from the Computer Service and became an independent department. This gave Margaret a new set of challenges as she had to establish those administrative services (from reception to health and safety and building services) that had previously been supplied by the Computing Service. The fruition of her labours was evident at the time of the building's official opening on May 1st 2002.

Margaret has enjoyed being part of the department. "I've been very fortunate in being part of the Lab during a period of such change and growth!" Her enjoyment stems in large part from the ethos fostered by Roger Needham: that the department should be very democratic with everyone allowed their say.

So, with Margaret part of the very fibre of the Computer Laboratory, there will be something missing when she leaves before the start of the new academic year. However, we expect to see her regularly when she returns for the weekly Pilates and aerobics classes. And, rest assured, Margaret will be just as busy in retirement as she has been during her career. She plans to spend more time playing the piano, painting, gardening, and decorating — there may even be time for a part-time job!

Whatever Margaret does in her retirement we all wish her a very happy one, and thank her heartily for leaving the Computer Laboratory in such robust health.

J. L. (Larry) Nazareth

Numerical Algorist vis-à-vis Numerical Analyst

Dedicated to the memory of **Richard E. Crandall** (1947–2012)

Few would argue with the observation that mathematics in full flower as we know it today, both pure and applied, has evolved from the root concept of number. This is recounted beautifully in the book *Number: The Language of Science* by Tobias Dantzig, which was first published in 1930 and then appeared in several subsequent editions. (The author's son, George B. Dantzig, the inventor of linear programming and the simplex algorithm, achieved even greater eminence than his father.) This landmark book was endorsed by Albert Einstein as follows:

> "This is beyond doubt the most interesting book on the evolution of mathematics that has ever fallen into my hands. If people know how to treasure the truly good, this book will attain a lasting place in the literature of the world. The evolution of mathematical thought from the earliest times to the latest constructions is presented here with admirable consistency and originality and in a wonderfully lively style."

Nowadays every schoolchild learns *number representation* at an early age, along with the basic arithmetic operations on decimal *numerals*. But the *concept of number* itself is far from elementary, a fact highlighted by the great mathematician D. E. Littlewood, Fellow of Trinity College, in the chapter "Numbers" of his classic primer, *A Skeleton Key of Mathematics* [1947; 2002]:

> "A necessary preliminary for any proper understanding of mathematics is to have a clear conception of what is meant by number. When dealing with number most people refer to their own past handling of numbers, and this is, usually, not inconsiderable. Familiarity gives confidence in the handling, but not always an insight into the significance. The technique of manipulating numbers is learned by boys and girls at a very tender age when manipulative skill is fairly easily obtained, and when the understanding is very immature. At a later stage, when the faculty of understanding develops, the technique is already fully acquired, so that it is not necessary to give any thought to numbers. To appreciate the significance of numbers it is necessary to go back and reconsider the ground which was covered in childhood. Apart from specialized mathematicians, few people realize that, for example, the number 2 can have half a dozen distinct meanings. These differences in meaning are reflected in the logical definitions of number."

Littlewood then proceeds to explain these "differences in meaning" and gives a brief yet masterful exposition of the logical foundations of four basic number systems: cardinal integers; signed integers; rational numbers; and real numbers. And this line of development could be continued, whereby the underlying *structure* of the foregoing number systems is generalised, extended or relaxed, leading to other key developments in mathematics: vector spaces, matrix and tensor theory, groups, rings and fields, functional analysis, and so on.

In contrast, computer science is a much younger discipline, although it too has roots that stretch back to antiquity. Its key foundational concepts are *algorithm* and *universal machine*, and these foundations were laid during the 1930s — well before the advent of the electronic computer — by a group of mathematical logicians, including Gödel, Church, Kleene, Post and, above all, Turing. (Useful background information on these pioneers and their contributions can be found in Berlinski [2000].) Seemingly different formulations of "algorithm" were shown to be equivalent, leading to what became known later within computer science as "*the grand unified theory of computation*." In the public discourse, a "recipe" in a cookbook is often used as an analogue for "algorithm". But, in reality, recipe (say within a soup cookbook) stands in relation to algorithm in the same way that numeral stands in relation to number. Like number, the concept of "algorithm" is far from elementary: the analogue of an algorithm is, in fact, closer to the *entire* soup cookbook, with different choices of ingredients (inputs to an algorithm) leading to different soups (outputs). Another frequently-used term in this setting is "program", the concrete realisation of an algorithm as a finite list of instructions. Unlike the latter, a program — itself a closer analogue to recipe — is not required to produce an answer for each and every input. In other words, every algorithm gives rise to a program within a prescribed model of computation, or given computer language, but every program is *not* the realisation of an algorithm; hence the ubiquitous, so-called halting problem, a key breakthrough of Turing, which can be stated very simply as follows: within a given model of computation, does there exist a particular (universal) algorithm that can examine any program whatsoever within the model and determine whether or not that program is an algorithm?

Academic departments of computer science themselves only came into existence within universities during the 1960s — in a few instances 1950s. One of the first was created by the great computer pioneer Maurice Wilkes at Cambridge University, where it was known initially, during my student days, as the Mathematical Laboratory. Numerical analysts played a key role in setting up such departments and, in particular, they initiated the process of bringing *number under the rubric of algorithm*. However, by making the primary focus of attention the finite-precision, floating-point number system, they achieved this objective in only a very limited, albeit practically important, way. The works of Jim Wilkinson, a close collaborator of Turing, were the landmark achievements in this area (see Wilkinson [1963], [1965]).

In a contribution on numerical analysis to the Princeton Companion to Mathematics (Gowers *et al.* [2008]), Nick Trefethen of Oxford University summarises this state of affairs as follows:

> "In the 1950s and 1960s, the founding fathers of the field *[of numerical analysis]* discovered that inexact arithmetic can be a source of danger, causing errors in results that "ought" to be right. The source of such problems is numerical instability, that is, the amplification of rounding errors from microscopic to macroscopic scale by certain modes of computation. These men, including Von Neumann, Wilkinson, Forsythe, and Henrici, took pains to publicise the risks of careless reliance on computer arithmetic. These risks are very real, but the message was communicated all too successfully, leading to the widespread impression that the main business of numerical analysis is coping with rounding errors. In fact, the main business of numerical analysis is designing algorithms that converge quickly; rounding error analysis, while a part of the discussion, is rarely the central issue. If rounding error vanishes, 90% of numerical analysis would remain.
>
> Numerical analysis sprang from mathematics; then it spawned the field of computer science. When universities began to found computer science departments in the 1960s, numerical analysts were often in the lead. Now, two generations later, most of them are to be found in mathematics departments. What happened? A part of the answer is that numerical analysts deal with continuous mathematical problems, whereas computer scientists prefer discrete ones, and it is remarkable how wide a gap that can be.

At the time of this re-migration of numerical analysts back to mathematics during the 1970s and 80s, other key developments occurred within theoretical computer science, in particular, breakthroughs by Stephen Cook and Richard Karp on NP-completeness, and the identification of the all-encompassing P=NP problem of computational complexity. Later, during the 1990s, the study of models of computation and complexity within computer science — the so–called "grand unified theory" — leapfrogged back into mathematics, whence the subject had originated, thanks largely to the work of the Fields Medalist Stephen Smale and his co-workers. This provided a theoretical foundation for numerical analysis within mathematics and is described by Smale as follows (quoted from the panel discussion in Renegar et al. [1996]):

> "A lot of my motivation in spending so much time trying to understand numerical analysis is to help my own ideas about how to define an algorithm. It seems to me that it is important [if one is] to understand the subject of numerical analysis to make a definition of algorithm. It is the main object of study of numerical analysis and to have a definition of it so someone can look at all algorithms or a class of algorithms is an important line of understanding."

And he adds:

> "…numerical analysis does not need these things. It doesn't need a model of computation. But on the other hand, I think that [it] will develop. It's going to develop anyway, and it is going to develop probably more in parallel with existing analysis numerical. Numerical analysis will do very fine without it. But in the long run, these ideas from geometry and foundations will give a lot of insights and understanding of numerical analysis."

In their resulting landmark work, Blum, Cucker, Shub, and Smale [1998] presented a computational model (BCSS) of great generality — abstract machines defined over mathematical rings and fields — and then developed a theory of computational complexity, in particular over the real and complex numbers. One could summarise this activity within the field of mathematics as setting out *to bring algorithm under the rubric of number*, the former being appropriately delineated as in the foregoing reference, and the latter being broadly conceived as the branches of mathematics that flowed from the number concept.

A compelling argument can be made in favour of undertaking *a converse and complementary activity within computer science*, namely, seeking to bring number under the rubric of algorithm, the former again being appropriately delineated and the latter broadly conceived as the aforementioned "grand unified theory of computation". As previously noted, this activity was initiated by numerical analysts during the 1960s, but it was left largely unfinished after their repatriation to mathematics. In a recently published monograph titled *Numerical Algorithmic Science and Engineering: Foundations and Organization* (PSIPress, Portland, Oregon, USA, 2012, xx+166 pgs.), I've sought to make some headway on this complementary activity. I make a distinction between *symbol*-based and *magnitude*-based models of computation; introduce a real-number model of computation that *re–conceptualises the floating-point number system* so as to be able to address foundational issues and not just practical ones; compare and contrast the role of the *numerical algorist* (an introduced term used to identify a specialist in numerical algorithmic science and engineering within computer science) with that of the *numerical analyst* (a specialist in numerical analysis within mathematics); note that the two can co-exist peacefully and indeed complement one another; observe that numerical algorithmic science and engineering (AS&E) addresses discrete-numerical (as contrasted with combinatorial) and continuous, finite-dimensional problems, and that the watershed between numerical AS&E and numerical analysis is identified more by the partition between finite-dimensional and infinite-dimensional than by the partition between discrete and continuous (with the two disciplines overlapping in the continuous, finite-dimensional domain). And so on…

The book, a manifesto of sorts, was published by PSIPress, a new venture in science and technology publishing. PSIPress was founded recently by Richard E. Crandall, a close and long-term associate of the late Steve Jobs of Apple Computers. Sadly, Crandall was also taken ill and died suddenly and unexpectedly last year, just before Christmas; see the blogs of Wolfram [2012] and Borwein and Bailey [2012] for obituaries/tributes to this most remarkable polymath.

The complete front-matter of my book — cover images, preface, full table of contents, and a detailed introduction — along with the first chapter can be viewed at the publisher's Web site www.psipress.com, where a very inexpensive PDF version of the book can be purchased and downloaded directly, using a choice of payment options. A limited number of inexpensive paperback copies is also available. I believe the book will be of interest to readers of *The Ring* and I would very much welcome feedback, which can be sent to me at either of the following e-mail addresses: nazareth@amath.washington.edu or larry_nazareth@q.com. I would be glad to respond and engage in a dialogue on this subject.

References

Berlinski, D. (2000), The Advent of the Algorithm, Harcourt, N.Y.

Blum, L., et al. (1998), Complexity and Real Computation, Springer-Verlag, N.Y.

Borwein, J. and Bailey, D. (2012), http://experimentalmath.info/blog/2012/12/mathematicianphysicistinventor-richard-crandall-dies-at-64

Dantzig, T. (1954), Number, Free Press, Macmillan, N.Y., 4th Edition.

Gowers, T. et al. (2008), The Princeton Companion to Mathematics, PUP, Princeton, NJ.

Littlewood, D. E. (2002), The Skeleton Key of Mathematics, Dover, N.Y.

Nazareth, J. L. (2012), Numerical Algorithmic Science and Engineering, PSIPress, Oregon.

Renegar, J. et al. (1996), Mathematics of Numerical Analysis, AMS, Providence, RI

Wilkinson, J. H.. (1963), Rounding Errors in Algebraic Processes, Prentice-Hall, NJ

Wilkinson, J. H.. (1965), The Algebraic Eigenvalue Problem, Clarendon Press, Oxford.

Wolfram, S. (2012), http://blog.stephenwolfram.com/2012/12/remembering-richard-crandall-1947-2012/

Author's Bio

John Lawrence (Larry) Nazareth is Professor Emeritus in the Department of Mathematics at Washington State University (USA) and Affiliate Professor in the Department of Applied Mathematics at the University of Washington. He was educated at the University of Cambridge (Trinity, '64) where he obtained a bachelor's degree in mathematics in 1967 and a postgraduate diploma in computer science at the Mathematical Laboratory — today's Computer Laboratory — in 1968. He then moved to the United States and continued his graduate studies at the University of California, Berkeley, specialising in the areas of optimisation and numerical computation. In 1973, he completed his doctorate in computer science and simultaneously obtained a master's degree in operations research. He is the author of numerous technical articles and he has written seven books within the field of numerical algorithmic science and engineering. He currently resides on Bainbridge Island in the Seattle area, Washington, USA. For further background information, visit http://www.math.wsu.edu/faculty/nazareth .

In the author's words: "I became part of the general drift, described in this article, of people involved with numerical computation moving from computer science back to mathematics. But my basic orientation has always remained that of the computer scientist. The main thrust of my book on numerical AS&E is that this drift, now decades in the making, should be arrested and reversed, because substantial benefits are to be gained from computer science re-embracing the "numerical algorist". This book marks a return to my roots in computer science, harking back to my time at the Cambridge Mathematical Laboratory, when the mathematically meaningless equation i=i+1 first became a FORTRAN metaphor for the wonder-filled world of algorithms."

Hall of Fame news

Bango

Bango has been placed on the 2013 FinTech 50 Watchlist, a shortlist of 50 companies in Europe that are redefining financial technology.

Bango has reached 100 mobile operator connections globally, 200 million billable identities and several marquee partnerships including Facebook, Microsoft and Telefónica Digital. Providing mobile payment for most of the world's largest app stores, Bango is emerging as the de-facto standard for payments on the mobile Web.

Cronto

comdirect bank AG has followed Commerzbank AG in deploying CrontoSign (known in Germany as photoTAN), Cronto's visual transaction signing solution. CrontoSign is a simple and effective way to mitigate even very sophisticated attacks by Trojan malware.

FusePump

FusePump, a provider of multi-channel e-commerce solutions for on-line retail, has been selected by ModelZone to power its products into on-line marketplaces and affiliate channels. FusePump's marketplaces integration solution will allow ModelZone, the UK's largest specialist model chain, to power its products into Amazon, eBay and Play.com.

Global Inkjet Systems

Global Inkjet Systems (GIS), a leading developer of software drivers and electronics for industrial inkjet printheads, has received a Queen's Award for Enterprise 2013 in the International Trade category. This prestigious award recognises the company's outstanding achievement in growing revenues from overseas markets in Europe, USA and Asia.

GIS technology is used in a wide range of inkjet applications including labels, packaging, ceramic tiles, product decoration, textiles and 3D printing systems.

Health2Works

Rally Round, a Web-based service that helps families to give more practical help to frail older relatives, has launched a new app.

Rally Round is one of ten featured apps on the Department of Health's new NHS Apps library that was launched at NHS Expo 2013. It can also be found on the Apple App store and at www.rallyroundme.com.

Jagex

Jagex, UK's biggest computer games developer, made a pre-tax profit of £9.8m as revenues climbed to £53m.

Jagex is 55% owned by Insight Venture Partners following its purchase of co-founder Andrew Gower's 33% stake for £75m.

Linguamatics

Linguamatics has been included in Outsells' *Information Industry Outlook*. As part of *Outlook*, Outsell named its annual "30 to Watch" — companies that shake up their marketplaces by innovating and challenging the industry status quo.

Mango Health

Mango Health has released the newest version of Mango Health on the Apple App Store.

Masabi

Masabi, the leader in transit mobile ticketing and agile fare collection, has secured US$2.8m in funding from Detroit-based Fontinalis Partners, London-based MMC Ventures and existing investor m8 Capital, to accelerate US transit ticketing. It is also opening an office in New York.

Netronome

Netronome has raised US$19m in equity capital from four venture capital firms.

The money from Sourcefire, Intel Capital, DFJ Esprit and the Raptor Group will enable Netronome to accelerate research and development to meet demand for its latest product, called NFP-6xxx.

Raspberry Pi

Raspberry Pi has released the US$30 add-on camera board. The camera it is a 5MP Omnivision 5647 sensor in a fixed-focus module, typical of the kinds of units seen in some mid–range camera phones.

The Raspberry Pi Foundation has appointed Clive Beale as Director of Educational Development. Mr Beale previously worked at Kesgrave High School where he taught ICT and Computing.

RealVNC

RealVNC has been singled out with its third Queen's Award for Enterprise in as many years, a remarkable achievement for the provider of VNC® remote access and control software. This year the company has been recognised for its outstanding achievements in International Trade, following its unique double win for sustained achievements in Innovation and International Trade in 2011.

RealVNC has rapidly expanded its overseas sales, with exports comprising over 90% of turnover and overseas sales growth for the three-year period assessed at over 250%.

RealVNC has a diverse, international customer base with VNC users spread across over 175 countries.

Sintefex Audio

Satronen Sound chose JoeCo BlackBox Recorders, designed in partnership with Sintefex, to capture Mumford & Sons' *Gentlemen of the Road* tour.

Spektrix

The Tron Theatre in Glasgow and New Wolsey Theatre in Ipswich have deployed Spektrix, the Web-based box office software for the arts.

SwiftKey

SwiftKey has unveiled SwiftKey Tilt. SwiftKey Tilt makes it possible to text without even touching the screen: just rock, wiggle or shimmy your smartphone to insert words! SwiftKey Tilt works by unleashing a pinball into the keyboard to power a third way to type on your device. While tapping or flowing words, the device sends the brightly coloured ball across the keyboard and when it collides with a prediction, the word is inserted. This frees up thumbs to make even quicker progress through a text, email or Tweet. SwiftKey assures customers that it is fully compatible with both the Macarena and the Harlem Shake!

Events calendar

2013

June

Tuesday 4th, 6.30pm
The Ring Summer Garden Party
Henry J Bean's, 195–197 King's Road, Chelsea, London SW3 5ED

August

Thursday 1st, 6.30pm
London Ringlet Bar
Venue to be confirmed

October

Wednesday 3rd, 6.30pm
Venue to be confirmed

December

Tuesday 3rd, 6.30pm
Venue to be confirmed

3D Constrained Local Model for Rigid and Non-Rigid Facial Tracking

Tadas Baltrušaitis Peter Robinson
University of Cambridge Computer Laboratory
15 JJ Thomson Avenue
tb346@cl.cam.ac.uk pr10@cl.cam.ac.uk

Louis-Philippe Morency
USC Institute for Creative Technologies
12015 Waterfront Drive
morency@ict.usc.edu

Abstract

We present 3D Constrained Local Model (CLM-Z) for robust facial feature tracking under varying pose. Our approach integrates both depth and intensity information in a common framework. We show the benefit of our CLM-Z method in both accuracy and convergence rates over regular CLM formulation through experiments on publicly available datasets. Additionally, we demonstrate a way to combine a rigid head pose tracker with CLM-Z that benefits rigid head tracking. We show better performance than the current state-of-the-art approaches in head pose tracking with our extension of the generalised adaptive view-based appearance model (GAVAM).

1. Introduction

Facial expression and head pose are rich sources of information which provide an important communication channel for human interaction. Humans use them to reveal intent, display affection, express emotion, and help regulate turn-taking during conversation [1, 12]. Automated tracking and analysis of such visual cues would greatly benefit human computer interaction [22, 31]. A crucial initial step in many affect sensing, face recognition, and human behaviour understanding systems is the estimation of head pose and detection of certain facial feature points such as eyebrows, corners of eyes, and lips. Tracking these points of interest allows us to analyse their structure and motion, and helps with registration for appearance based analysis. This is an interesting and still an unsolved problem in computer vision. Current approaches still struggle in person-independent landmark detection and in the presence of large pose and lighting variations.

There have been many attempts of varying success at tackling this problem, one of the most promising being the Constrained Local Model (CLM) proposed by Cristinacce and Cootes [10], and various extensions that followed [18, 23, 27]. Recent advances in CLM fitting and response functions have shown good results in terms of accuracy and convergence rates in the task of person independent facial feature tracking. However, they still struggle in under poor lighting conditions.

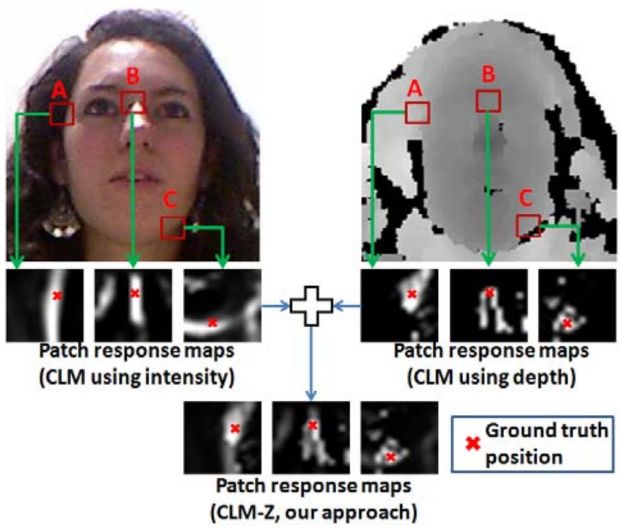


Figure 1. Response maps of three patch experts: (A) face outline, (B) nose ridge and (C) part of chin. Logistic regressor response maps [23, 27] using intensity contain strong responses along the edges, making it hard to find the actual feature position. By integrating response maps from both intensity and depth images, our CLM-Z approach mitigates the aperture problem.

In this paper, we present a 3D Constrained Local Model (CLM-Z) that takes full advantage of both depth and intensity information to detect facial features in images and track them across video sequences. The use of depth data allows our approach to mitigate the effect of lighting conditions. In addition, it allows us to reduce the effects of the aperture problem (see Figure 1), which arises because of patch response being strong along the edges but not across them. An additional advantage of our method is the option to use depth only CLM responses when no intensity signal is available or lighting conditions are inadequate.

Furthermore, we propose a new tracking paradigm which integrates rigid and non-rigid facial tracking. This paradigm integrates our CLM-Z with generalised adaptive view-based appearance model (GAVAM) [19], leading to better head pose estimation accuracy. We make the code, landmark labels and trained models available for research purposes[1].

We evaluate our approaches on four publicly available datasets: the Binghamton University 3D dynamic facial expression database (BU-4DFE) [30], the Biwi Kinect head pose database (Biwi) [14], the Boston University head pose database (BU) [6], and our new dataset ICT-3DHP. The experiments show that our method significantly outperforms existing state-of-the-art approaches both for person-independent facial feature tracking (convergence and accuracy) and head pose estimation accuracy.

First, we present a brief overview of work done in facial feature point and head pose tracking (Section 2). Then we move on to formulate the CLM-Z problem and describe the fitting and model training used to solve it (Section 3). Additionally, we present an approach to rigid-pose tracking that benefits from non-rigid tracking (Section 3.4). Finally we demonstrate the advantages of our approaches through numerical experiments (Section 4).

2. Related work

Non-rigid face tracking refers to locating certain landmarks of interest from an image, for example nose tip, corners of the eyes, and outline of the lips. There have been numerous approaches exploring the tracking and analysis of such facial feature points from single images or image sequences [16, 21, 31].

Model-based approaches show good results for feature point tracking [16]. Such approaches include Active Shape Models [9], Active Appearance Models [7], 3D Morphable Models [2], and Constrained Local Models [10].

Feature points in the image are modelled using a point distribution model (PDM) that consists of non-rigid shape and rigid global transformation parameters. Once the model is trained on labelled examples (usually through combination of Procrustes analysis and principal component analysis), a fitting process is used to estimate rigid and non-rigid parameters that could have produced the appearance of a face in an unseen 2D image. The parameters are optimised with respect to an error term that depends on how well the parameters are modelling the appearance of a given image, or how well the current points represent an aligned model.

Constrained Local Model (CLM) [10] techniques use the same PDM formulation. However, they do not model the appearance of the whole face but rather the appearance of local patches around landmarks of interest (and are thus similar to Active Shape Model approaches). This leads to more generalisability because there is no need to model the complex appearance of the whole face. The fitting strategies employed in CLMs vary from general optimisation ones to custom tailored ones. For a detailed discussion of various fitting strategies please refer to Saragih *et al.* [23].

There are few approaches that attempt tracking feature points directly from depth data, most researches use manually labeled feature points for further expression analysis [15]. Some notable exceptions are attempts of deformable model fitting on depth images directly through the use of iterative closest point like algorithms [3, 5]. Breidt *et al.* [3] use only depth information to fit an identity and expression 3D morphable model. Cai *et al.* [5] use the intensity to guide their 3D deformable model fitting. Another noteworthy example is that of Weise *et al.* [28], where a person-specific deformable model is fit to depth and texture streams for performance based animation. The novelty of our work is the full integration of both intensity and depth images used for CLM-Z fitting.

Rigid head pose tracking attempts to estimate the location and orientation of the head. These techniques can be grouped based on the type of data they work on: *static*, *dynamic* or *hybrid*. Static methods attempt to determine the head pose from a single intensity or depth image, while dynamic ones estimate the object motion from one frame to another. Static methods are more robust while dynamic ones show better overall accuracy, but are prone to failure during longer tracking sequences due to accumulation of error [20]. Hybrid approaches attempt to combine the benefits of both static and dynamic tracking.

Recent work also uses depth for static head pose detection [4, 13, 14]. These approaches are promising, as methods that rely solely on 2D images are sensitive to illumination changes. However, they could still benefit from additional temporal information. An approach that uses intensity and can take in depth information as an additional cue, and combines static and dynamic information was presented by Morency *et al.* [19] and is described in Section 3.4.

Rigid and non-rigid face tracking approaches combine head pose estimation together with feature point tracking. There have been several extensions to Active Appearance Models that explicitly model the 3D shape in the formulation of the PDM [29], or train several types of models for different view points [8].They show better performance for feature tracking at various poses, but still suffer from low accuracy at estimating the head pose.

Instead of estimating the head pose directly from feature points, our approach uses a rigid-pose tracker that is aided by a non-rigid one for a more accurate estimate.

3. CLM-Z

The main contribution of our paper is CLM-Z, a Constrained Local Model formulation which incorporates intensity and depth information for facial feature point tracking.

Our CLM-Z model can be described by parameters

[1] http://www.cl.cam.ac.uk/research/rainbow/emotions/

$\mathbf{p} = [s, \mathbf{R}, \mathbf{q}, \mathbf{t}]$ that can be varied to acquire various instances of the model: the scale factor s, object rotation $\mathbf{R}$ (first two rows of a 3D rotation matrix), 2D translation $\mathbf{t}$, and a vector describing non-rigid variation of the shape $\mathbf{q}$. The point distribution model (PDM) used in CLM-Z is:

$$\mathbf{x}_i = s \cdot \mathbf{R}(\overline{\mathbf{x}}_\mathbf{i} + \mathbf{\Phi}_i \mathbf{q}) + \mathbf{t}. \quad (1)$$

Here $\mathbf{x}_i = (x, y)$ denotes the 2D location of the i^{th} feature point in an image, $\overline{\mathbf{x}}_i = (X, Y, Z)$ is the mean value of the i^{th} element of the PDM in the 3D reference frame, and the vector $\mathbf{\Phi}_i$ is the i^{th} eigenvector obtained from the training set that describes the linear variations of non-rigid shape of this feature point.

This formulation uses a weak-perspective (scaled orthographic) camera model instead of perspective projection, as the linearity allows for easier optimisation. The scaling factor s can be seen as the inverse of average depth and the translation vector $\mathbf{t}$ as the central point in a weak-perspective model. This is a reasonable approximation due to the relatively small variations of depth along the face plane with respect to the distance to camera.

In CLM-Z we estimate the maximum *a posteriori* probability (MAP) of the face model parameters $\mathbf{p}$ in the following equation:

$$p(\mathbf{p}|\{l_i\!=\!1\}_{i=1}^n, \mathcal{I}, \mathcal{Z}) \propto p(\mathbf{p}) \prod_{i=1}^{n} p(l_i\!=\!1|\mathbf{x}_i, \mathcal{I}, \mathcal{Z}), \quad (2)$$

where $l_i \in \{1, -1\}$ is a discrete random variable indicating if the i^{th} feature point is aligned or misaligned, $p(\mathbf{p})$ is the prior probability of the model parameters $\mathbf{p}$, and $\prod_{i=1}^n p(l_i = 1|\mathbf{x}_i, \mathcal{I}, \mathcal{Z})$ is the joint probability of the feature points $\mathbf{x}$ being aligned at a particular point $\mathbf{x}_i$, given an intensity image $\mathcal{I}$ and a depth one $\mathcal{Z}$.

Patch experts are used to calculate $p(l_i = 1|\mathbf{x}_i, \mathcal{I}, \mathcal{Z})$, which is the probability of a certain feature being aligned at point $\mathbf{x}_i$ (from Equation 1).

3.1. Patch experts

We estimate if the current feature point locations are aligned through the use of local patch experts that quantify the probability of alignment ($p(l_i = 1|\mathbf{x}_i, \mathcal{I}, \mathcal{Z})$) based on the surrounding support region.

As a probabilistic patch expert we use Equation 3; the mean value of two logistic regressors (Equations 4, and 5).

$$p(l_i|\mathbf{x}_i, \mathcal{I}, \mathcal{Z}) = 0.5 \times (p(l_i|\mathbf{x}_i, \mathcal{I}) + p(l_i|\mathbf{x}_i, \mathcal{Z})) \quad (3)$$

$$p(l_i|\mathbf{x}_i, \mathcal{I}) = \frac{1}{1 + e^{d\mathcal{C}_{\mathcal{I},i}(\mathbf{x}_i;\mathcal{I}) + c}} \quad (4)$$

$$p(l_i|\mathbf{x}_i, \mathcal{Z}) = \frac{1}{1 + e^{d\mathcal{C}_{\mathcal{Z},i}(\mathbf{x}_i;\mathcal{Z}) + c}} \quad (5)$$

Here $\mathcal{C}_{\mathcal{Z},i}$ and $\mathcal{C}_{\mathcal{I},i}$ are the outputs of intensity and depth patch classifiers, respectively, for the i^{th} feature, c is the logistic regressor intercept, and d the regression coefficient.

We use linear SVMs as proposed by Wang *et al.* [27], because of their computational simplicity, and efficient implementation on images using convolution. The classifiers can thus be expressed as:

$$\mathcal{C}_{\mathcal{I},i}(\mathbf{x}_i; \mathcal{I}) = \mathbf{w}_{\mathcal{I},i}^T \mathcal{P}_{\mathcal{I}}(\mathcal{W}(\mathbf{x}_i; \mathcal{I})) + b_{\mathcal{I},i}, \quad (6)$$

$$\mathcal{C}_{\mathcal{Z},i}(\mathbf{x}_i; \mathcal{Z}) = \mathbf{w}_{\mathcal{Z},i}^T \mathcal{P}_{\mathcal{Z}}(\mathcal{W}(\mathbf{x}_i; \mathcal{Z})) + b_{\mathcal{Z},i}, \quad (7)$$

where $\{\mathbf{w}_i, b_i\}$ are the weights and biases associated with a particular SVM. Here $\mathcal{W}(\mathbf{x}_i; \mathcal{I})$ is a vectorised version of $n \times n$ image patch centered around $\mathbf{x}_i$.

$\mathcal{P}_{\mathcal{I}}$ normalises the vectorised patch to zero mean and unit variance. Because of potential missing data caused by occlusions, reflections, and background elimination we do not use $\mathcal{P}_{\mathcal{I}}$ on depth data, we use a robust $\mathcal{P}_{\mathcal{Z}}$ instead. Using $\mathcal{P}_{\mathcal{I}}$ on depth data, missing values skew the normalised patch (especially around the face outline) and lead to bad performance (see Figures 3, 4).

$\mathcal{P}_{\mathcal{Z}}$ ignores missing values in the patch when calculating the mean. It then subtracts that mean from the patch and sets the missing values to an experimentally determined value (in our case 50mm). Finally, the resulting patch is normalised to unit variance.

Example images of intensity, depth and combined response maps (the patch expert function evaluated around the pixels of an initial estimate) can be seen in Figure 1. A major issue that CLMs face is the aperture problem, where detection confidence across the edge is better than along it, which is especially apparent for nose ridge and face outline in the case of intensity response maps. Addition of the depth information helps with solving this problem, as the strong edges in both images do not correspond exactly, providing further disambiguation for points along strong edges.

3.2. Fitting

We employ a common two step CLM fitting strategy [10, 18, 23, 27]; performing an exhaustive local search around the current estimate of feature points leading to a response map around every feature point, and then iteratively updating the model parameters to maximise Equation 2 until a convergence metric is reached. For fitting we use Regularised Landmark Mean-Shift (RLMS) [23].

As a prior $p(\mathbf{p})$ for parameters $\mathbf{p}$, we assume that the non-rigid shape parameters $\mathbf{q}$ vary according to a Gaussian distribution with the variance of the i^{th} parameter corresponding to the eigenvalue of the i^{th} mode of non-rigid deformation; the rigid parameters s, $\mathbf{R}$, and $\mathbf{t}$ follow a non-informative uniform distribution.

Treating the locations of the true landmarks as hidden variables, they can be marginalised out of the likelihood that the landmarks are aligned:

$$p(l_i|\mathbf{x}_i, \mathcal{I}, \mathcal{Z}) = \sum_{\mathbf{y}_i \in \mathbf{\Psi}_i} p(l_i|\mathbf{y}_i, \mathcal{I}, \mathcal{Z}) p(\mathbf{y}_i|\mathbf{x}_i), \quad (8)$$

where $p(\mathbf{y}_i|\mathbf{x}_i) = \mathcal{N}(\mathbf{y}_i; \mathbf{x}_i, \rho\mathbf{I})$, with ρ denoting the variance of the noise on landmark locations arising due to PCA truncation in PDM construction [23], and $\mathbf{\Psi}_i$ denotes all integer locations within the patch region.

By substituting Equation 8 into Equation 2 we get:

$$p(\mathbf{p}) \prod_{i=1}^{n} \sum_{\mathbf{y}_i \in \Psi_i} p(l_i|\mathbf{y}_i, \mathcal{I}, \mathcal{Z}) \mathcal{N}(\mathbf{y}_i; \mathbf{x}_i, \rho\mathbf{I}). \quad (9)$$

The MAP term in Equation 9 can be maximised using Expectation Maximisation [23].

Our modification to the original RLMS algorithm is in the calculation of response maps and their combination. Our new RLMS fitting is as follows:

Algorithm 1 Our modified CLM-Z RLMS algorithm

Require: $\mathcal{I}$, $\mathcal{Z}$ and $\mathbf{p}$
  Compute intensity responses { Equation 4 }
  Compute depth responses { Equation 5 }
  Combine the responses {Equation 3}
  while not converged($\mathbf{p}$) **do**
    Linearise shape model
    Compute mean-shift vectors
    Compute PDM parameter update
    Update parameters
  end while
  return $\mathbf{p}$

We use Saragih *et al.*'s [23] freely available implementation of RLMS[2]. The difference between the available implementation and the algorithm described in Saragih *et al.* [23], is through the use of patches trained using profile face images in addition to frontal ones. This leads to three sets of classifiers (frontal, left, right), with the left and right sets not having the response functions for the occluded landmarks. This enables us to deal with self occlusion as the invisible points are not evaluated for the fitting procedure.

3.3. Training

Training CLM-Z involves constructing the PDM and training the patch experts. The point distribution model is used to both provide the prior $p(\mathbf{p})$ and to linearise the shape model. The patch experts serve to calculate the response maps.

We use the PDM provided by Saragih *et al.* [23], which was created using non-rigid structure from motion [24] approach on the Multi-PIE [17] dataset.

For the intensity-based SVM classifiers and the logistic regressors, we used the classifiers used by Wang *et al.* [27] and Saragih *et al.* [23]. The local descriptors were trained on the Multi-PIE [17] dataset using 400 positive and $15k$ negative examples for each landmark for frontal images, and 30 positive examples for profile images, due to the lack of labeled data. The interocular distance of the training images was 30 pixels, and the patch sizes used for training were 11×11 pixels.

Currently there is no extensive dataset with labeled facial feature points of depth images over varying poses. Collecting such a dataset would be very time consuming and costly, especially if a wide range of poses is to be covered; manually labelling feature points on depth images would also be very difficult (see depth images in Figure 2).

In order to create such a training set we use the 4D-BUFE [30] as our starting point. 4D-BUFE consists of video sequences of 101 subjects acting out one of the six basic emotions. It was collected using the Di3D[3] dynamic face capturing system, which records sequences of texture images together with 3D models of faces. This means that by labelling the feature points in the texture images we are able to map them to the 3D models of faces. The 3D models can then be rotated and rendered at various poses. This allows us to generate many labelled depth images from a single labelled texture image.

We took a subset of 707 frames (each participant with neutral expression and peaks of the 6 basic emotions) and labelled the images with 66 feature points semi-automatically (with the aid of the intensity based CLM tracker followed by manual inspection and correction). The original 3D models were rotated from $-70°$ to $70°$ yaw, and $-30°$ to $30°$ pitch and their combinations. Examples of the rendered training data can be seen in Figure 2.

We trained the depth-based classifiers using 400 positive and $15k$ negative examples for each feature for every experiment (making sure that subject independence is preserved). The interocular distance and patch sizes were the same as for intensity training data.

3.4. Combining rigid and non-rigid tracking

Because non-rigid shape based approaches, such as CLM, do not provide an accurate pose estimate on their own (see Section 4.2), we present a way our CLM-Z tracker can interact with an existing rigid pose tracker. For a rigid head pose tracker we use a Generalised Adaptive View-based Appearance Model (GAVAM) introduced by Morency *et al.* [19]. The tracker works on image sequences and estimates the translation and orientation of the head in three dimensions with respect to the camera in addition to providing an uncertainty associated with each estimate.

GAVAM is an adaptive keyframe based differential tracker. It uses 3D scene flow [25] to estimate the motion of the frame from keyframes. The keyframes are collected and adapted using a Kalman filter throughout the video stream.

[2] http://web.mac.com/jsaragih/FaceTracker/FaceTracker.html (accessed Apr. 2012)

[3] http://www.di3d.com (accessed Apr. 2012)

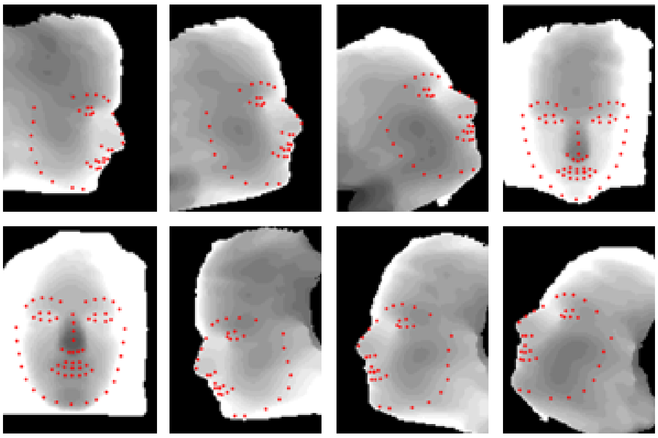


Figure 2. Examples of synthetic depth images used for training. Closer pixels are darker, and black is missing data.

This leads to good accuracy tracking and limited drift. The tracker works on both intensity and depth video streams. It is also capable of working without depth information by approximating the head using an ellipsoid. We introduce three extensions to GAVAM in order to combine rigid and non-rigid tracking, hence improving pose estimation accuracy both in the 2D and 3D cases.

Firstly, we replace the simple ellipsoid model used in 2D tracking with a person specific triangular mesh. The mesh is constructed from the first frame of the tracking sequence using the 3D PDM of the fitted CLM. Since different projection is assumed by CLM (weak-perspective) and GAVAM (full perspective), to convert from the CLM landmark positions to GAVAM reference frame we use:

$$Z_g = \frac{1}{s} + Z_p, X_g = Z_g \frac{x_i - c_x}{f}, Y_g = Z_g \frac{y_i - c_y}{f}, \quad (10)$$

where f is the camera focal length, c_x, c_y the camera central points, s is the PDM scaling factor (inverse average depth for the weak perspective model), Z_p the Z component of a feature point in PDM reference frame x_i, y_i the feature points in image plane, and X_g, Y_g, Z_g the vertex locations in the GAVAM frame of reference.

Secondly, we use the CLM tracker to provide a better estimate of initial head pose than is provided by the static head pose detector used in GAVAM. Furthermore, the initial estimate of head distance from the camera used in GAVAM (assumption that the head is 20 cm wide), is replaced with a more stable assumption of interpupillary distance of 62 mm [11], based on the tracked eye corners using the CLM-Z or CLM trackers.

Lastly, we provide additional hypotheses using the current head pose estimate from CLM-Z (CLM in 2D case) to aid the GAVAM tracker with the selection of keyframes to be used for differential tracking.

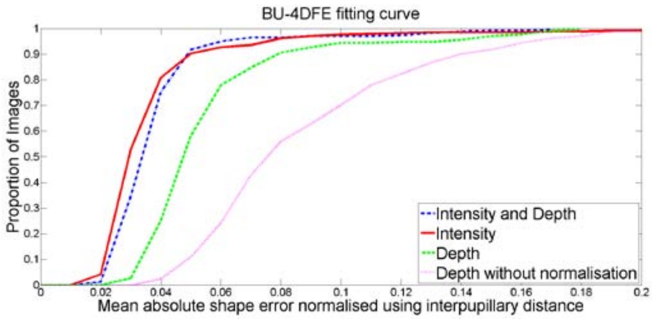


Figure 3. The fitting curve of CLM on intensity and depth images separately on the BU-4DFE dataset. Note the higher fitting accuracy on depth images using our normalisation scheme $\mathcal{P}_\mathcal{Z}$, as opposed to zero mean unit variance one

4. Experiments

To validate our CLM-Z approach and the extensions made to the rigid-pose tracker we performed both rigid and non-rigid tracking experiments that demonstrate the benefits of our methods. In the following section when we refer to CLM we mean the CLM formulation presented by Saragih *et al.* [23] which uses RLMS for fitting, and linear SVMs with logistic regressors as patch experts.

4.1. Non-rigid face tracking

4.1.1 BU-4DFE

For this experiment we split the data into two subsets: training and testing. Training set included 31 female and 20 male subjects, while the testing 26 female and 24 male subjects. We discarded some images from the training and test sets due to lack of coverage by the range scanner (e.g. part of the chin is missing in the range scan). This lead to 324 3D models used for generating the training data (see Section 3.3), and 339 texture and depth images for testing. The average Inter-ocular distance of the resulting test set was 300 pixels.

The tracker was initialised by an off the shelf Viola-Jones [26] face detector. The fit was performed using 11×11 pixel patch experts on a 15×15 pixel search window. The error was measured by using the mean absolute distance from the ground truth location for each feature point.

You can see the comparison of intensity and depth signals in Figure 3. Intensity modality manages to track the feature points better than the depth one. However, the depth modality on its own is still able to track the feature points well, demonstrating the usefuleness of depth when there is no intensity information available. We can also see the benefit of using our specialised normalisation $\mathcal{P}_\mathcal{Z}$. The small difference in intensity and intensity with depth tracking is because the original CLM is already able to track the faces in this dataset well (frontal images with clear illumination), and the advantage of adding depth is small.

Method	Converged	Mean error
CLM intensity	64 %	0.135
CLM depth with $\mathcal{P}_{\mathcal{Z}}$	50%	0.152
CLM depth without $\mathcal{P}_{\mathcal{Z}}$	13%	0.16
CLM-Z	**79%**	**0.125**

Table 1. Results of feature point tracking on Biwi dataset. Measured in absolute pixel error. The mean errors are reported only for the converged frames (<0.3 of interocular distance)

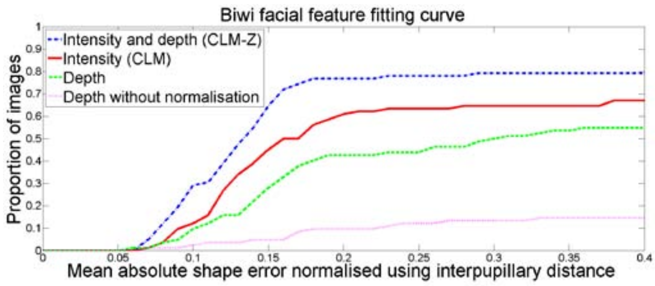


Figure 4. The fitting curve of CLM and CLM-Z on the Biwi dataset facial feature point subset. Note that intensity and depth combined lead to best performance. Furthermore, depth without $\mathcal{P}_{\mathcal{Z}}$ normalisation fails to track the videos succesfully.

4.1.2 Biwi

There currently is no facial feature point labelled video sequence dataset that contains depth information, thus we chose to use a publicly available head pose dataset and label a subset of it with feature points.

We used the Biwi Kinect head pose dataset [14]. It consists of 24 video sequences collected using the Microsoft Kinect sensor. For this experiment we selected 4 videos sequences of 772, 572, 395, and 634 frames each. We manually labeled every 30th frame of those sequences with 66 feature points (or in the case of profile images 37 feature points), leading to 82 labeled images in total. This is a particularly challenging dataset for a feature point tracker due to large head pose variations ($\pm 75°$ yaw and $\pm 60°$ pitch).

The training and fitting strategies used were the same as for the previous experiment. For feature tracking in a sequence the model parameters from the previous frame were used as starting parameters for tracking the next frame. We did not use any reinitialisation policy because we wanted to compare the robustness of using different patch responses in CLM fitting, and a reinitialisation policy would have influenced some of the results.

The results of this experiment can be seen in Table 1 and Figure 4. We see a marked improvement of using our CLM-Z over any of the modalities separately (depth or intensity). Furthermore, even though using only depth is not as accurate as using intensity or combination of both it is still able to track the sequences making it especially useful under very bad lighting conditions where the standard CLM tracker is prone to failure. Furthermore, we see the benefit of our normalisation function $\mathcal{P}_{\mathcal{Z}}$.

Even though the training and testing datasets were quite different (high resolution range scanner was used to create the training set and low resolution noisy Kinect data for testing) our approach still managed to generalise well and improve the performance of a regular CLM without any explicit modeling of noise. The examples of tracks using CLM and CLM-Z on the Biwi dataset can be seen in Figure 5.

Method	Yaw	Pitch	Roll	Mean
Regression forests [14]	7.17	9.40	7.53	8.03
GAVAM [19]	3.00	3.50	3.50	3.34
CLM [23]	11.10	9.92	7.30	9.44
CLM-Z	6.90	7.06	10.48	8.15
CLM-Z with GAVAM	**2.90**	**3.14**	**3.17**	**3.07**

Table 2. Head pose estimation results on ICT-3DHP. Error is measured in mean absolute distance from the ground truth.

4.2. Rigid head pose tracking

To measure the performance of our rigid pose tracker we evaluated it on three publicly available datasets with existing ground truth head pose data. For comparison, we report the results of using Random Regression Forests [13] (using the implementation provided by the authors), and the original GAVAM implementation.

4.2.1 ICT-3DHP

We collected a head pose dataset using the Kinect sensor. The dataset contains 10 video sequences (both intensity and depth), of around 1400 frames each and is publicly available[4]. The ground truth was labeled using a Polhemus FASTRAK flock of birds tracker. Examples of tracks using CLM and CLM-Z on our dataset can be seen in Figure 6.

Results of evaluating our tracker on ICT-3DHP can be seen in Table 2. We see a substantial improvement of using GAVAM with CLM-Z over all other trackers.

From the results we see that a CLM and CLM-Z trackers are fairly inaccurate for large out of plane head pose estimation, making them not very suitable for human head gesture analysis on their own. However, the inaccuracy in roll when using CLM, and CLM-Z might be explained by lack of training data images displaying roll.

4.2.2 Biwi dataset

We also evaluated our approach on the dataset collected by Fanelli *et al.* [14]. The dataset was collected with a frame based algorithm in mind so it has numerous occasions of

[4] http://projects.ict.usc.edu/3dhp/

Method	Yaw	Pitch	Roll	Mean
Regression forests [14]	9.2	8.5	**8.0**	8.6
CLM	28.85	18.30	28.49	25.21
CLM-Z	14.80	12.03	23.26	16.69
CLM-Z with GAVAM	**6.29**	**5.10**	11.29	**7.56**

Table 3. Head pose estimation results on the Biwi Kinect head pose dataset. Measured in mean absolute error.

Method	Yaw	Pitch	Roll	Mean
GAVAM [19]	3.79	4.45	2.15	3.47
CLM [23]	5.23	4.46	2.55	4.08
CLM with GAVAM	**3.00**	**3.81**	**2.08**	**2.97**

Table 4. Head pose estimation results on the BU dataset. Measured in mean absolute error.

lost frames and occasional mismatch between colour and depth frames. This makes the dataset especially difficult for tracking based algorithms like ours whilst not affecting the approach proposed by Fanelli *et al*. [13]. Despite these shortcomings we see an improvement of tracking performance when using our CLM-Z with GAVAM approach over that of Fanelli *et al*. [13] (Table 3).

4.2.3 BU dataset

To evaluate our extension to the 2D GAVAM tracker we used BU dataset presented by La Cascia *et al*. [6]. It contains 45 video sequences from 5 different people with 200 frames each. The results of our approach can be seen in Table 4. Our approach significantly outperforms the GAVAM method in all of the orientation dimensions.

5. Conclusion

In this paper we presented CLM-Z, a Constrained Local Model approach that fully integrates depth information alongside intensity for facial feature point tracking. This approach was evaluated on publicly available datasets and shows better performance both in terms of convergence and accuracy for feature point tracking from a single image and in a video sequence. This is especially relevant due to recent availability of cheap consumer depth sensors that can be used to improve existing computer vision techniques.

Using only non-rigid trackers for head pose estimation leads to less accurate results than using rigid head pose trackers. Hence, we extend an existing rigid-pose GAVAM tracker to be able to use the non-rigid tracker information leading to more accuracy when tracking head pose.

In future work we will explore the possibility of using a prior for rigid transformation parameters from GAVAM instead of a uniform distribution that is currently used in CLM and CLM-Z. We would also like to explore the use of a perspective camera model for CLM-Z fitting. This will lead to more integration between rigid and non-rigid trackers.

In addition, we will explore the use of different classifiers for patch experts, as what is appropriate for intensity image might not be suitable for depth information. Moreover, we would like to explore the influence of noise for the CLM-Z fitting, as the training data used was clean which is not the case for the consumer depth cameras.

Acknowledgments

We would like to thank Jason Saragih for suggestions and sharing his code which served as the inspiration for the paper. We would also like to thank Lech Świrski and Christian Richardt for helpful discussions. We acknowledge funding support from Thales Research and Technology (UK). This material is based upon work supported by the U.S. Army Research, Development, and Engineering Command (RDECOM). The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

[1] N. Ambady and R. Rosenthal. Thin Slices of Expressive behavior as Predictors of Interpersonal Consequences : a Meta-Analysis. *Psychological Bulletin*, 111(2):256–274, 1992. 1

[2] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *SIGGRAPH*, pages 187–194, 1999. 2

[3] M. Breidt, H. Biilthoff, and C. Curio. Robust semantic analysis by synthesis of 3d facial motion. In *FG*, pages 713 –719, march 2011. 2

[4] M. Breitenstein, D. Kuettel, T. Weise, L. van Gool, and H. Pfister. Real-time face pose estimation from single range images. In *CVPR*, pages 1 –8, june 2008. 2

[5] Q. Cai, D. Gallup, C. Zhang, and Z. Zhang. 3d deformable face tracking with a commodity depth camera. In *ECCV*, pages 229–242, 2010. 2

[6] M. L. Cascia, S. Sclaroff, and V. Athitsos. Fast, Reliable Head Tracking under Varying Illumination: An Approach Based on Registration of Texture-Mapped 3D Models. *TPAMI*, 22(4):322–336, 2000. 2, 7

[7] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *TPAMI*, 23(6):681–685, Jun 2001. 2

[8] T. Cootes, K. Walker, and C. Taylor. View-based active appearance models. In *FG*, pages 227–232, 2000. 2

[9] T. F. Cootes and C. J. Taylor. Active shape models - 'smart snakes'. In *BMVC*, 1992. 2

[10] D. Cristinacce and T. Cootes. Feature detection and tracking with constrained local models. In *BMVC*, 2006. 1, 2, 3

[11] N. A. Dodgson. Variation and extrema of human interpupillary distance, in stereoscopic displays and virtual reality systems. In *Proc. SPIE 5291*, pages 36–46, 2004. 5

[12] P. Ekman, W. V. Friesen, and P. Ellsworth. *Emotion in the Human Face*. Cambridge University Press, second edition, 1982. 1

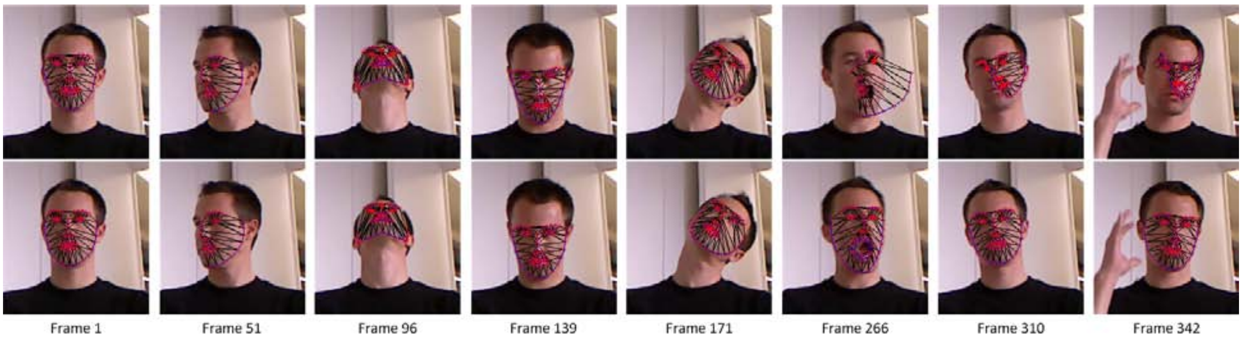


Figure 5. Examples of facial expression tracking on Biwi dataset. Top row CLM, bottom row CLM-Z.

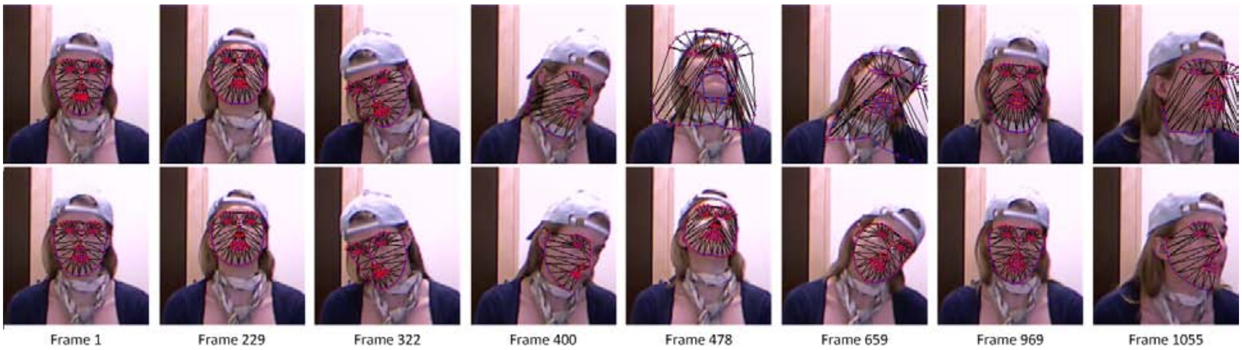


Figure 6. Examples of facial expression tracking on our dataset. Top row CLM, bottom row our CLM-Z approach.

[13] G. Fanelli, J. Gall, and L. V. Gool. Real time head pose estimation with random regression forests. In *CVPR*, pages 617–624, 2011. 2, 6, 7

[14] G. Fanelli, T. Weise, J. Gall, and L. van Gool. Real time head pose estimation from consumer depth cameras. In *DAGM*, 2011. 2, 6, 7

[15] T. Fang, X. Zhao, O. Ocegueda, S. Shah, and I. Kakadiaris. 3D facial expression recognition: A perspective on promises and challenges. In *FG*, pages 603 –610, 2011. 2

[16] X. Gao, Y. Su, X. Li, and D. Tao. A review of active appearance models. *Trans. Sys. Man Cyber. Part C*, 40(2):145–158, 2010. 2

[17] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. *IVC*, 28(5):807 – 813, 2010. 4

[18] L. Gu and T. Kanade. A generative shape regularization model for robust face alignment. *ECCV*, pages 413–426, 2008. 1, 3

[19] L.-P. Morency, J. Whitehill, and J. R. Movellan. Generalized Adaptive View-based Appearance Model: Integrated Framework for Monocular Head Pose Estimation. In *FG*, pages 1–8, 2008. 2, 4, 6, 7

[20] E. Murphy-Chutorian and M. M. Trivedi. Head Pose Estimation in Computer Vision: A Survey. *TPAMI*, 31:607–626, 2009. 2

[21] M. Pantic and M. Bartlett. *Machine Analysis of Facial Expressions*, pages 377–416. I-Tech Education and Publishing, 2007. 2

[22] P. Robinson and R. el Kaliouby. Computation of emotions in man and machines. *Phil. Trans. of the Royal Society B: Biological Sciences*, 364(1535):3441–3447, 2009. 1

[23] J. Saragih, S. Lucey, and J. Cohn. Deformable Model Fitting by Regularized Landmark Mean-Shift. *Int. J. Comput. Vision*, 91(2):200–215, Jan. 2011. 1, 2, 3, 4, 5, 6, 7

[24] L. Torresani, A. Hertzmann, and C. Bregler. Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors. *TPAMI*, 30(5):878 –892, may 2008. 4

[25] S. Vedula, S. Baker, P. Rander, R. T. Collins, and T. Kanade. Three-dimensional scene flow. In *ICCV*, pages 722–729, 1999. 4

[26] P. Viola and M. J. Jones. Robust real-time face detection. *Int. J. Comput. Vision*, 57(2):137–154, 2004. 5

[27] Y. Wang, S. Lucey, and J. Cohn. Enforcing convexity for improved alignment with constrained local models. In *CVPR*, June 2008. 1, 3, 4

[28] T. Weise, S. Bouaziz, H. Li, and M. Pauly. Realtime performance-based facial animation. *SIGGRAPH*, 30(4), 2011. 2

[29] J. Xiao, S. Baker, I. Matthews, and T. Kanade. Real-time combined 2d+3d active appearance models. In *CVPR*, volume 2, pages 535 – 542, June 2004. 2

[30] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale. A high-resolution 3d dynamic facial expression database. In *FG*, pages 1–6, 2008. 2, 4

[31] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *TPAMI*, 31(1):39–58, 2009. 1, 2