Application

2015-03-31

2015-04588	**Li, Haibo**	**NT-2**

Information about applicant

Name: Haibo Li

Doctorial degree: 1993-11-15

Birthdate: 19651008

Academic title: Professor

Gender: Male

Employer: Kungliga Tekniska högskolan

Administrating organisation: Kungliga Tekniska högskolan

Project site: MID, Medieteknik och interaktionsdesign

Information about application

Call name: Forskningsbidrag Stora utlysningen 2015 (Naturvetenskap och teknikvetenskap)

Type of grant: Projektbidrag

Focus: Fri

Subject area:

Project title (english): Vision-Based Hand Articulation Tracking for "Whole-Hand" Human-Computer Interaction

Project start: 2016-01-01

Project end: 2019-12-31

Review panel applied for: NT-2, NT-14, NT-1

Classification code: 10209. Medieteknik

Keywords: HCI, Hand gesture, Computer Vision

Funds applied for

Year:	2016	2017	2018	2019
Amount:	1,446,000	1,512,000	1,610,000	1,733,000

Descriptive data

Project info

Project title (Swedish)*

Vision-baserat handen artikulation spårning för "Hela-Hand" människa-dator interaktion

Project title (English)*

Vision-Based Hand Articulation Tracking for “Whole-Hand” Human-Computer Interaction

Abstract (English)*

With the introduction of more smart mobile devices, iphone, ipad, iwatch, and VR glass which augments virtual worlds over our physical world, we need gesture technologies to support us to interact naturally with virtual (digital) objects with our hands, that is, manipulation type interaction. Combined with the dexterity and hand–eye coordination skills of humans, the hand can be turned into an easy to use control device with high degrees of freedom. Our research question in this project is how to turn human hand into a whole-hand input device technically for human computer interaction.

Inspired by how Google solves web-size text search problem, we will develop a new framework for tracking articulated hand motions from vision based on search technologies. Our innovation is to re-define the problem of hand articulation tracking as a text retrieval problem. The idea is to build a huge database containing millions of hand posture images. These posture images should “emulate” all kinematically possible articulated hand motions. These millions images can be pre-analyzed and tagged with 3D hand motion parameters including joint angles of articulated fingers. When the hand of a user is captured by visual sensors, the captured hand image is used to retrieve the most similar hand posture image stored in the database. Then the motion parameters tagged with the matched image are given to the captured hand image. Thus, hand articulation tracking is achieved. To quickly find the best matched image from a database, our solution is to treat each hand image as a document, a hand shape as a “sentence”, convert the shape into “word” and “phrase”, and employ the powerful text retrieval tool, inverted indexing, to do the search. Since the computing of whole hand articulation is done offline, millions and hundred millions sets of articulation parameters could be evaluated in advance and results can be pre-downloaded into devices. This guarantees the effectiveness and efficiency of the solution. We believe that the framework gives us a totally fresh looking to the classic computer vision problem.

Popular scientific description (Swedish)*

Den snabba utvecklingen av mobila enheter under det senaste decenniet har i stor utsträckning drivits av interaktion och visualiseringsteknologi. Även om pekskärmar avsevärt har förbättrat interaktions tekniken är det förutsägbart att med framtida mobila enheter, t.ex. augmented reality glasögon och smarta klockor, kommer användare kräver mer intuitiva sätt att interagera, såsom ex. fri hand interaktion i 3D-rymden. Speciellt viktigt blir det vid manipulation av digitalt innehåll i utökade miljöer där 3D hand/kropp gester kommer att vara ytterst nödvändig. Därför är 3D gest igenkänning och spårning högt önskade egenskaper för interaktionsdesign i framtida smarta miljöer. På grund av komplexiteten i hand/kroppsrörelser, och begränsningar av mobila enheter vid dyra beräkningar, är 3D-gest analys fortfarande ett mycket svårt problem att lösa.

Projektet syftar till att införa nya begrepp, teorier och tekniker för naturlig och intuitiv interaktion i framtida utökade miljöer. Bidrag från denna projekt stöder begreppet naken-hand 3D gest interaktion och interaktiv visualisering på framtida smarta enheter. De införda tekniska lösningar möjliggör effektiv interaktion i 3D-rymden runt den smarta enheten. Hög noggrannhet och robust 3D rörelseanalys av hand/kropp gester utförs för att underlätta 3D-interaktion i olika tillämpningsscenarier. De föreslagna teknik gör det möjligt för användare att kontrollera, manipulera, och organiserar digitalt innehåll i 3D-rymden.

Project period

Number of project years*

4

Calculated project time*

2016-01-01 - 2019-12-31

Classifications

Select a minimum of one and a maximum of three SCB-codes in order of priority.

Select the SCB-code in three levels and then click the lower plus-button to save your selection.

SCB-codes*

1. Naturvetenskap > 102. Data- och informationsvetenskap (Datateknik) > 10209. Medieteknik

---

Enter a minimum of three, and up to five, short keywords that describe your project.

Keyword 1*

HCI

Keyword 2*

Hand gesture

Keyword 3*

Computer Vision

Keyword 4

Keyword 5

Ethical considerations

Specify any ethical issues that the project (or equivalent) raises, and describe how they will be addressed in your research. Also indicate the specific considerations that might be relevant to your application.

Reporting of ethical considerations*

In this project human hands will be used for pattern recognition. We will ask one or two users to perform all sorts of gestures with their bare-hands against a uniform background. These hand gestures are coming from natural manipulating interactions. No any harmful issues are associated with the usage of the hands.

Hand gestures are recorded and used for converting into binary hand shape images. Motion sensors are attached to the hands. The ground true hand motion and articulation parameters will be obtained and tagged to the hand posture images. However, we don't store any hand shape images. Only abstract shape words and the associated articulation parameters are stored and made in a table. Furthermore, no personal identity is recorded. The abstract words and parameters have not any connection with a person. Therefore, there is no privacy issue involved.

The project includes handling of personal data

Yes

The project includes animal experiments

No

Account of experiments on humans

No

Research plan

Vision-Based Hand Articulation Tracking for “Whole-Hand” Human-Computer Interaction

1. Purpose and Aims

Gesture-based interfaces have dramatically increased in popularity over the past few years. Microsoft's Kinect is a good example of a controller-free interface where hand and arm movements as well as postural information are being tracked as input to a computing system. The major advantage with gesture-based interfaces is that they allow users to focus on the task and not on the input device. Furthermore, the embodied nature of gesture interaction tends to be more engaging. Due to technical challenges, hand gestures have just been used for *command* and *control* type interaction, for example, around “point”, “reach”, and “grab” interaction or signing. Although this addresses some need for natural user interface, it doesn’t explore the full capabilities and dexterity of the hand as an input device.

With the introduction of smart mobile devices, iphone, ipad, iwatch, and VR glass which augments virtual worlds over our physical world, we need gesture technologies to support us to interact naturally with virtual (digital) objects with our hands, that is, *manipulation* type interaction. Combined with the dexterity and hand–eye coordination skills of humans, the hand can be turned into an easy to use control device with high degrees of freedom (DOF). Our **research question** in this project is *how to turn hand(s) into a whole-hand(s) input device technically for human computer interaction.* “Whole-hand input” is a well defined gesture technology in the HCI field[2]. The distinguishing characteristic of whole-hand input is that the user does not think in terms of manipulating an input device as is the case with other haptic forms of input (e.g. mouse, joystick, trackball), but moves his hand to directly affect the task. The salient features of using the whole hand as an input device can be divided into three principal categories: **naturalness**, **adaptability**, and **dexterity**. Naturalness means users can use pre-acquired sensorimotor skills for different types of grips, specific finger coordination, while adaptability refers to the hand's ability to quickly and smoothly switch functions, and dexterity can be defined as the integration of movements and senses into higher levels of competence.

The major technology bottleneck behind turning hand into a whole-hand input device lies in the difficulty of capturing and analysing the articulated hand motions. Human hand is a highly articulated object with more than 27 degrees of freedom[11]. It is very challenging to extract these 27 articulation parameters from interacting hands in real-time. Existing solutions are to employ digital gloves (glove-type devices), which directly measure the finger joint angles and spatial positions of the hand by using a set of sensors (e.g. electromagnetic or fiber-optical sensors). Although they exist in applications of human computer interaction, virtual reality, and game fields, digital gloves are regarded too intrusive, cumbersome, and expensive for natural interactions, particularly with mobile devices. This calls for technical alternatives.

Computer vision (CV) has the potential to provide more natural, non-contact solutions. Computer vision has a distinctive role in the development of direct sensing-based HCI. However, various challenges must be addressed in order to satisfy the demands of potential interaction methods. Currently, vision-based hand gesture estimation has some limitations in processing arbitrary hand actions. Incorporating the full functionality of the hand in HCI requires capturing the whole hand articulated motion. Right now CV can provide support for only a small range of hand actions under restrictive conditions that mainly serve to issue commands. Other types of gestures such as gesticulations and manipulation operation, or high DOF interaction require more efficient general hand articulation estimation algorithms. Full DOF hand gesture estimation is still a big challenge in CV[1]. The state of the art algorithms[3] [4] are adopting a model-based approach where the 3D kinematic hand model is used to render hand poses to recover 27 articulation parameters. The high dimensionality of the problem and self-occlusions are the main obstacles. For example, the model-based approach requires finding a solution from 10^{20} possible parameter configurations in 10ms!

Inspired by how Google solves web-size text search problem, in this project we will develop a new **framework** for tracking articulated hand motions from vision based on search technologies. Our innovation is to **re-define** *the problem of hand articulation tracking as a text retrieval problem*. The **idea** is to build a huge database containing millions of hand posture images. These posture images should "emulate" all kinematically possible articulated hand motions. These millions images can be pre-analyzed and tagged with 3D hand motion parameters including joint angles of articulated fingers. When the hand of a user is captured by visual sensors, the captured hand image is used to retrieve the most similar hand posture image stored in the database. Then the motion parameters tagged with the matched image are given to the captured hand image. Thus, hand articulation tracking is achieved. The key of this approach is how to quickly find the best matched image from a database. Our **solution** is to treat each hand image as a document, a hand shape as a "sentence", convert the shape into "word" and "phrase", and employ the powerful text retrieval tool, inverted indexing, to do the search. Since the computing of whole hand articulation is done offline, millions and hundred millions sets of articulation parameters could be evaluated in advance and results can be pre-downloaded into devices. This guarantees the effectiveness and efficiency of the solution. We believe that the framework gives us a totally fresh looking to the classic computer vision problem.

Therefore, in this project we first put a **scientific hypothesis**: *Text retrieval approach is an effective and efficient solution to whole hand articulation tracking problem*. The research **objective** is to test the hypothesis by developing a theoretical framework of tracking articulated hand motions and evaluating with a proof of concept. The **aims** of this project are

- Increase general understanding and insight with respect to one of the hard problems in computer vision: *tracking articulated objects through sequences of images*, which is the core technology to enable the whole hand input for HCI.
- Introduce a theoretical framework of tracking articulated hand motions and develop associated core algorithms with a goal of leading to a new computing paradigm of solving hand articulation tracking problem.
- Work on a proof of concept to evaluate both effectiveness and efficiency of the proposed framework on both public and home created databases.

2. State of the Art

In this project we will NOT address hand gesture classification. Instead we focus on hand articulation tracking (full DOF hand pose estimation). Existing vision-based hand articulation tracking algorithms can be grouped into two categories[1]: *appearance based approaches* and *3D hand model based approaches*. Appearance-based approaches try to recover articulation parameters through detecting and locating salient hand and finger features, like fingertips, palm orientation, finger joins from 2D or 3D image (depth) features. The popular image features used for detecting these features include hand colours and shapes, local hand features, optical flow, 3D depth features, and so on. Our early work on hand tracking belongs to this type of approaches [7]. The drawback of these feature-based based approaches is that clean image segmentation is generally required in order to extract the hand features. This is not a trivial task when the background is cluttered, for instance. Furthermore, since human hands are highly articulated, it is often difficult to find local hand features due to self-occlusion, therefore, some kinds of heuristics are needed to handle the large variety of hand gestures. Instead of detecting and locating hand gesture features directly, 3D hand model based approaches use the 3D kinematic hand model to render hand poses. An analysis-by-synthesis (ABS) strategy is employed to recover the hand articulation motion parameters by matching the appearance projected by the 3D hand model with the observed posed hand image from the camera, and minimizing the discrepancy between them.

Generally speaking, it is easier to achieve real-time performance with appearance-based approaches due to that only simpler 2D image features are required. However, this type of approaches can only handle very limited number of simple hand gestures, like detection and tracking of fingertips, it is hard to achieve full DOF hand articulation tracking. In contrast, 3D hand model based approaches offer a rich description of hand articulation that potentially allows a full coverage of all possible hand gestures. The **bad** news is that a 3D hand model is controlled by 27 articulation parameters (27 DOF). The number of possible articulation parametric configurations is huge!

Saying that if each articulation parameter is quantized into 8 levels, thus the number of possible parameter configurations will be 10^{20} ($=8^{27}$). Furthermore, the hand has very fast motion capabilities with a speed reaching up to 5 m/s for translation and 300^{o}/s for wrist rotation, which requires 100 Hz frame rate for tracking. Obviously, one can be deeply skeptical about how technically possible to select a set of 27 articulation parameters from 10^{20} possible parameter combinations in 10 ms! The state of the art 3D model-based hand articulation tracking algorithms employ evolutionary optimization techniques, such as, the Particle Swarm Optimization (PSO) and quasi-random sampling to select very limited sets of articulation parameters, typically several hundreds, from huge parametric configurations to evaluate and determine the right set of 27 articulation parameters[3]. 10^{2} sets of configured parameters are very sparsely scattered in the huge 27-dimensional space. Evaluation only over 10^{2} sets of parameter configurations does not provide a creditable picture of 10^{20} sets of parameter configurations. This is why, to make the selected parameter configurations sense, these algorithms have to assume motion continuum and use the parameters from the previous frame as the start point for local search, and use pre-process, like search for Z-fingers to limit the scope of possible articulation parameters[4]. This will be getting worse when working on the case of two interacting hands, where one has to work in a 54 dimensional space and deal with 10^{40} possible parameter configurations! This is known as “curse of dimensionality”, which is a fatal obstacle for 3D model based approach to be a paradigm for real-time hand articulation tracking.

To recover full DOF articulation parameters in real-time we have to search for new paradigms. Recently, deep-learning and convolutional networks have been applied on the problem of hand articulation tracking from single depth images and shown some interesting results[5]. It is too early to judge if the deep-learning could be a new paradigm but definitely it is very strong in handle hand gesture classification and recognition problems. Another promising approach is to use text retrieval technology to search for articulation parameters. Zhou et al.

proposed an approach that integrates the powerful text retrieval tools with computer vision techniques in order to improve the efficiency for hand images retrieval [10]. An Okapi-Chamfer matching algorithm is used in their work based on the inverted index technique. Athitsos et al. proposed a method that can generate a ranked list of three-dimensional hand configurations that best match an input image [8]. Hand pose estimation is achieved by searching for the closest matches for an input hand image from a large database of synthetic hand images. The novelty of their system is the ability to handle the presence of clutter. Imai et al. proposed a 2D appearance-based method by using hand contours to estimate 3D hand posture [9]. In their method, the variations of possible hand contours around the registered typical appearances are trained from a number of computer graphic images generated from a 3D hand model. A low-dimensional embedded manifold is created to overcome the high computational cost of the large number of appearance variations.

Although the methods based on text retrieval are very promising, the relevant research works are too few (scattered in the literatures) to be visible in the field. The reason might be that the approach is too primary, or the results are not impressive due to small sizes of the databases used for experiments. Moreover, it might be also a consequence of the success of Kinect in real-time human body posture recognition and tracking: the statistical approaches (random forest tree, for example) adopted in Kinect are dominating the mainstream of hand gesture tracking. This effect is further enhanced by the introduction of several hardware devices based on depth sensors, for example, the one from the Leap Motion Company. This type of hand gesture sensor can ran at an interactive rate on consumer hardware and interact with moving objects in real-time. Despite of its fantastic demo, the Leap Motion sensor can in fact not handle full range of hand articulation. With regarding to the special requirements for real-time whole-hand tracking, we believe the most promising approach is to use text retrieval technologies for hand tracking.

3. Project Description

3.1 Research Strategy and Scientific Methodologies

Our research strategy is NOT to address general "whole-hand" input problem, instead to take *hand articulation tracking for mobile devices* as a **case study**. For next generation mobile devices, users will no longer satisfy with just performing interaction over 2D touch screen. They prefer to use whole hand as input for natural interaction. A "whole hand input interface" will be a must-have component in future mobile devices to facilitate the bare hands to manipulate digital (virtual) objects directly. Case study as a **scientific research method** allows us to have an up-close, in-depth, and detailed examination of the hand articulation tracking in real-life environments (contextual conditions). Working on a real-life case we could test the hypothesis through developing a new framework and algorithms and building a proof of concept to carry out both **qualitative** and **quantitative** studies. Considering its scientific, social and business significance, hand articulation tracking for mobile devices can be seen as a *critical* case having strategic importance in relation to the general hand articulation tracking problem.

3.2 The Definition of Research Problem

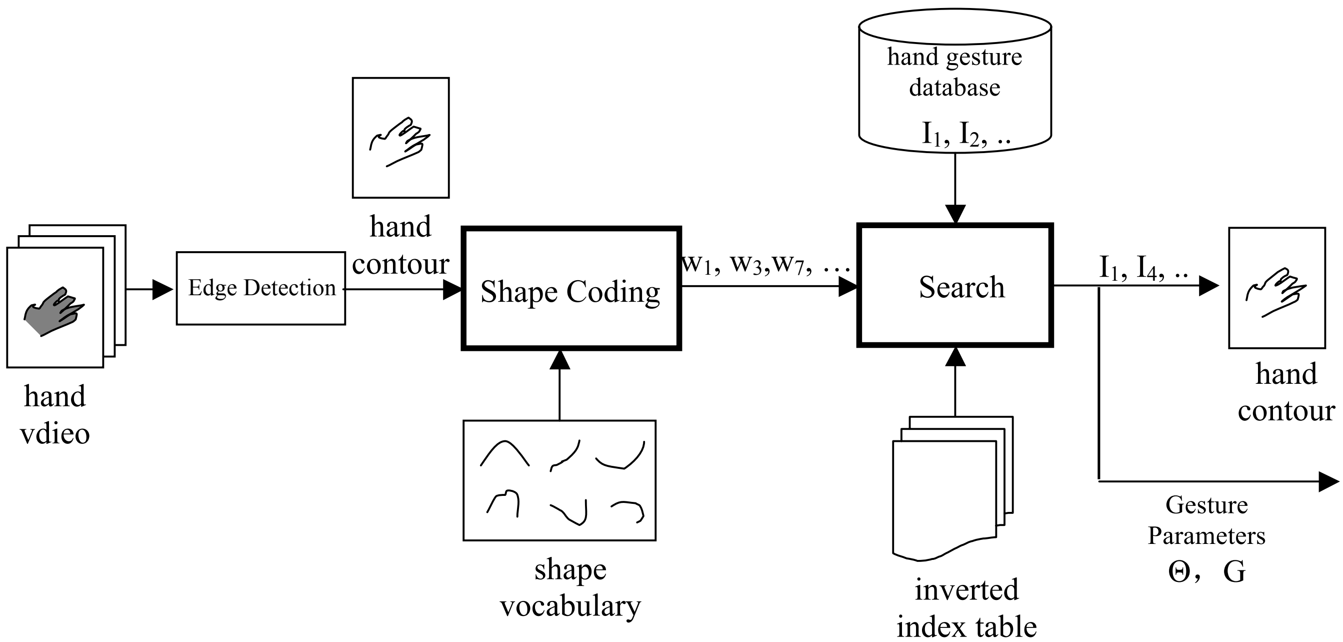
Our scientific approach is to *re-define the problem of hand articulation tracking as a text retrieval problem*. Targeting for real-world hand tracking applications, we have to address the hard problems of initialization and recovery also. Our strategy is to adopt the static approach, that is, *localizing and recognizing hand postures from individual frames* (here we assume 2D images as the visual observation. Technically, this is much more challenging than the case of depth images as observation. If depth sensors are available, the depth images can be viewed as 2D images. The same computing framework can be applied. Therefore, we will limit our study on 2D images as visual observation.

The framework is based on the idea of building a huge database which contains millions hand posture images, which in the best case emulate all possible hand articulated motions. Furthermore, these millions images are tagged with 3D hand motion parameters including joint angles of articulated fingers. When the hand of a mobile user is captured by the video camera on the back of the mobile phone, for example, the captured hand image is used to retrieve the most similar hand posture image stored in the database.

Here we give a mathematical definition of the problem: for a given hand contour image **Q**, we search for the most similar hand contour image **D** in a huge image database by computing the following distance function **E**:

$$(\theta^*, G^*) = \arg\min_{(\theta, G)} E(Q, D(\theta, G))$$

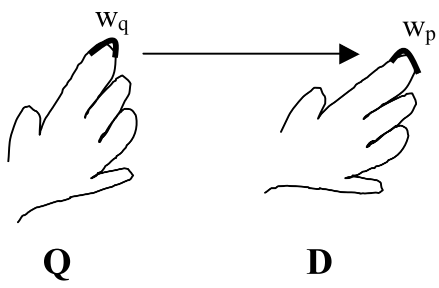
Here, (**θ**, **G**) is 3D hand motion and gestures associated with the hand posture in **D**. Parameter **G** = (**R**, **T**) is the global hand motion consisting of six rotational and translational motion parameters **R** and **T**. **θ** represents local finger articulations consisting of 21 finger joint parameters. In the database, all hand contour images were tagged with the corresponding 3D gesture hand motion and gesture (**θ**, **G**). The **D*** with the minimum distance **E** to **Q** is taken as the most similar hand contour image. The hand motion and gesture (**θ***, **G***) tagged with **D*** are assigned to **Q** as its articulation motion parameters. This process is shown in the figure.



3.3. Framework

All images are converted into binary shape images through standard edge detection and boundary detection operations. A hand posture image is represented by multiple contours, and each contour consists of edge segments, which are the visual shape words we will define in this project. Thus, a hand contour consists of consecutive words $\{\mathbf{w}_p\}$. For a given hand contour **Q**, searching for a similar hand contour **D** from the database by computing the distance function **E**(**Q**, **D**).

The Chamfer Matching can be used to compare two contours. Assume that $\mathbf{w}_p$ represents p^{th} edge segment, the Chamfer Distance from a database shape image **D** to the query shape image **Q** is defined as follows:



$$E(Q,D) = \frac{1}{|D|}\sum_{p\in D}\min_{q\in Q} \| w_p - w_q \|_2$$

where |**D**| is the number of segments in image **D**.

In order to reduce its complexity from $O(|\mathbf{D}|\times|\mathbf{Q}|)$ to $O(|\mathbf{D}|)$, a binary similarity map (called Hit map) will be employed to speedup the computation.

The value of the position $\mathbf{w}_p$ on the Hit map is given by the following Hit function[6]:

$$Hit_Q(p) = \begin{cases} 1, \quad \exists q \in Q \qquad (\| w_p - w_q \|_2 < r) \\ \qquad 0, \quad otherwise \end{cases}$$

where r is the tolerance radius.

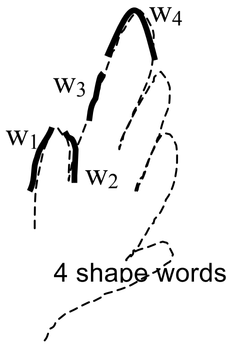
Thus, the similarity between two hand contour images **D** and **Q** can be calculated from the Hit map,

$$E(Q,D) = \frac{1}{|D|}\sum_{p\in D} Hit_Q(p)$$

Although it is not difficult to calculate **E**(**Q**, **D**), it is still quite time-consuming to compare a query hand contour **Q** with all millions hand images in the database: the complexity will be O(**K**), **K** is the total number of shape words containing in the database. For ten thousand shape words and millions database images, the computational complexity will be $O(10^{10})$, which is too heavy for mobile computing. Our approach is to build an inverted index structure to quickly perform matching from **Q** to **D** as Google does. Using such an inverted index structure, the computational complexity will be reduced to ○(**P**・**L**), where **P** is the average number of the non-zero elements in the Hit map and **L** is the average length of the inverted index table. Therefore, the design of shape vocabulary becomes critical. In this project we will develop a compact shape vocabulary where all shape words are of strong semantic contour shape. The design goal is the new shape vocabulary should enable the average length of the inverted index table in the range of $\mathbf{L} = 10^2$ and each hand contour image to contain $\mathbf{P} = 10^2$ words. Thus, the search complexity will be reduced to $○(10^4)$, which achieves a million-fold reduction in computational complexity. This is why search over millions images can be done in mobile devices in real-time.

3.4 Key Research Sub-Questions

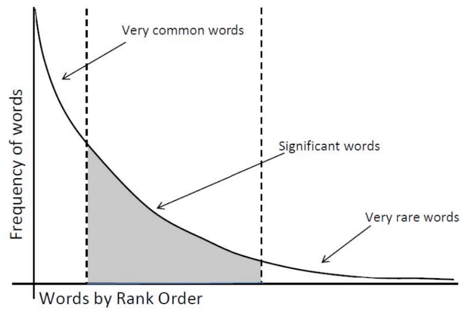
In the project we will concentrate on three fundamental research subquestions that we think are crucial for the success of the framework.



(1) How to Construct Effective and Efficient Shape Vocabulary?

The core in hand shape retrieval is how to represent shape contours. To formulate the hand articulation tracking problem into a text retrieval problem, two particular requirements are on the shape representation: 1) *shape sensitive*, the represented hand posture shape should be as close as possible to the one from input frame; 2) *position sensitive*, the matched posture shape should be at a similar position as the input hand posture.

It is critical to build good "shape vocabulary". By the "good" we mean the representation should be both effective and efficient. The idea is to divide a hand contour into segments. An individual segment is thought of as a word (see the figure on the right). To get an idea about the feasibility of this coding scheme, we give a simple analysis. Suppose that each contour segment, i.e. word, consists of 5 pixels. How many different word combinations could we have? Suppose that each pixel is allowed to connect only to the 3 neighbouring pixels in one direction. The number of word combinations for that direction in this case would be $1 \times 3^4 = 81$ words. For all four directions (right, left, up, and down), the total number of letters will be $4 \times 81 = 324$ words. However, most of these combinations would be complex, unnatural, and useless and thus represent noisy patterns. Our early research shows that 32 words are sufficient to represent the most hand shapes. The question is how to select the 32 words, which is related to another issue, how to assess the quality of the words. One way to do it is to look at the probability distribution of all words. To be able to perform efficient and effective search, shape vocabulary should follow the Zipf distribution. From the Zipf-type distribution, we will take away both the most common words and very rare words and just select the most important words (see figure right). This will greatly reduce the cost of storage and search time, making retrieval very efficient.



(2) How Large the Database Should Be? How to Build the Database?

Human hand is of a complex articulated structure with more than 27 DOF. To render all possible combinations of joints and poses huge numbers (10^{20}) of hand images will be generated. It is impossible to handle such a huge number of images. Fortunately, there is a strong correlation between finger joint angles. The state space of finger joints has a much low dimension. In [11] Wu *et al.* show that the state space for the joints can be approximated with 7 DOF. Thus, 7 is a rather good estimation of the embedded dimension of hand postures. If we quantize each DOF and represent it with 3 bits, thus we will have a total combination of 8^7= 2 millions states. Thus, the number of hand posture images stored in the database will be around 2 millions. We should stress that one advantage with such a representation is its capability of handling two interacting hands. If two hands are not interacting and we can use the single hand posture database to treat two hands individually. If two hands are interacting, we need another database to represent all possible interacting hand postures. The interaction between two hands greatly reduces the number of possible postures and a million hand postures should well cover all possible interacting hands.

The second question is how to build such a database. One solution is to use a 3D hand model to render all possible hand postures with computer graphics technology, and convert the generated hand posture images into binary shape images through edge detection and boundary detection. A potential problem with this approach is that the extracted edges are not natural-looking, which might affect shape match, and inherent dimensionality embedded in hand articulated motion is hard to emulated. Our innovative **solution** is to use a bare hand against a uniform background to perform all sorts of gestures. Hand gestures are recorded and used for converting into binary hand shape images. Motion sensors or video cameras are attached to the hand as we did in our early work. Thus, the ground true hand motion and articulation parameters can be obtained and tagged to the hand posture images.

(3) How to Handle Noisy Background?

Since the input hand image is captured from the real-world, the extracted hand contour will be very noisy. The noisy edges will give wrong index of shape words. This will lower the matching quality. To get satisfactory matching results we need a two-way matching, that is, the similarity between **Q** and **D** should be measured by performing matching both **D → Q** and **Q → D**. It is not straightforward to do the matching of **Q → D** since it is required to build distance transfer maps for all hand shape images in the database! Our strategy is to build shape “phrases” from 32 words and use the phrases to perform the **Q → D** matching quickly.

3.6 Timeline and Milestones

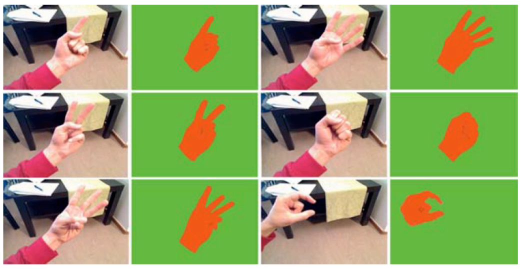
The project will stretch in four years and we expect to generate a PhD thesis. Here are the timeline and expected milestones

	Milestone
Year 1 **(WP1)**	*Construct Hand Posture Database* - record hand gesture images - compute 3D hand motion parameters of all gesture images.
Year 2 **(WP2)**	*Construct Shape Vocabulary* - collect training data - extract shape vocabulary by clustering techniques
Year 3 **(WP3)**	*Development of efficient search algorithms for matching shapes* - build an inverted index structure for D to Q matching. - develop a tree structure for Q to D matching.
Year 4 **(WP4)**	*Building a proof of concept to validate the framework* - build a proof of concept for evaluation and validation - evaluate the developed hand tracking algorithms - validate the new framework with real data - compare with model-based approach with public database

4. Scientific Significance

- Direct use of the whole hand as an input device will become next big thing for natural human–computer interaction (HCI). Currently, the only technology that satisfies the advanced requirements of hand-based input for HCI is glove-based sensing. This technology, however, has several drawbacks including that it hinders the ease and naturalness with which the user can interact with the computer-controlled environment and it requires long calibration and setup procedures. Computer vision (CV) has the highest potential to provide more natural, non-contact solutions.
- Articulated object tracking through sequences of images is one of the fundamental problems in computer vision. Although the introduction of 3D depth sensors has greatly lowered the difficulty of solving the problem, one still has to face with the inherent high-dimensionality problem arising from large number of DOF. Significant limitation from the model-based approach has called for an alternative technology. The approach based on the text retrieval framework has a high potential to be a new paradigm for solving the hard problem.
- The introduction of SIFT-types features has successfully solved quite many hard computer vision problems, including textured object recognition and tracking problem. However, there is still no any systematic framework on how to solve textureless objects (like human hands) recognition and tracking problem. It is of scientific significance to explore new types of visual features to handle edges or object contours. Viewing shape contours as text and characterizing them in three textual levels, "sentence", "phrase" and "word" is a new way to define visual features. This kind of shape representation has a potential to be part of new framework for handling hard computer vision problems associated with textureless objects.

5. Preliminary Results



In my PhD study I had proposed the principle, analysis-by-synthesis (ABS) for 3D model-based facial motion (global head motion plus facial expressions) analysis[12]. Today, the ABS principle has been used in hand articulation tracking. Although model-based approach has been used in computer vision for more than 20 years there are still no effective and efficient solutions to handle the inherent high dimensional parametric problem. Our research work on hand articulation tracking started from 2008. Two PhD students had been working in this topic. Both appearance-based and model-based approaches had been studied for whole hand input. The strategy of using search for matching hand shapes was popping out from the research process of when we had a hard time in using 3D model-based approach for real-time hand articulation tracking. In the second half of his PhD study the student sketched the basic framework of how to turn the tracking problem into a text retrieval problem. We built a prototype and run preliminary experiments. The experiments turned out surprised good results, 100% correct matching rate with a database containing more than 6000 gesture images[7]! This database has been quite challenging for model-based approach. 100% does not mean we solve the problem rather shows search-based approach is very promising. Today, we have extended this database to hold around a million hand postures, which will be used for our quantitative study. The database from Microsoft Research can be used for us to compare search based and model-based approach in a quantitative way.

Our research results in hand gesture recognition and tracking have generated two PhD theses, **two** Lic theses and **21** scientific publications. Some of them are listed here.

- Shahrouz Yousefi, Farid Abedan Kondori, Haibo Li, Camera-based Gesture Tracking For Real 3D Interaction Behind Mobile Devices, **The International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)**, Vol. 26, No. 8, 2012.

- Farid Abedan Kondori, Shahrouz Yousefi, Haibo Li, Direct 3D head pose estimation from Kinect-type sensors. **Electronics Letters**, Vol. 50, No. 4, pp. 268-270. 2013.

- Shahrouz Yousefi, Farid Abedan Kondori, Haibo Li, Experiencing Real 3D Gestural Interaction with Mobile Devices, **Pattern Recognition Letters**, Vol. 34, No. 8, 2013, pp. 912-921 , 2013,

- Zhihan Lu, A. Halawani, Haibo Li, "Hand and Foot Gesture Interaction for Handheld Devices", **ACM Transactions on Multimedia Computing Communications and Applications**. 10/2014; 11:1-19.

- Shahrouz Yousefi, Haibo Li, 3D Hand Gesture Analysis through a Real-time Gesture Search Engine， **International Journal of Advanced Robotic Systems**， 2015.

6. Key References

1. Erol, etc., "Vision-based Hand Pose Estimation: A Review", CVIU, 108(1-2), Oct. 2007.
2. D. Sturman, "Whole hand input", Ph.D. thesis, MIT, 1992
3. Oikonomidis, etc."Evolutionary quasi-random search for hand articulations tracking", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2014.
4. Qian, etc."Real-time and robust hand tracking from depth", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2014.
5. J. Tompson, etc, "Real-time continuous pose recovery of human hands using convolutional networks". In: TOG. (2014)
6. Yang Cao, etc, "Edgel index for large-scale sketch-based image search" IEEE Conference on Computer Vision and Pattern Recognition, June, 2011
7. S. Yousefi, "3D Gesture Recognition and Tracking for Next Generation of Smart Devices", PhD Thesis, KTH, 2014.
8. V. Athitsos and S. Sclaroff, "Estimating 3D hand pose from a cluttered image," in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003.
9. Imai, etc, "3-D hand posture recognition by training contour variation," in Proc. 6th IEEE International Conference on Automatic Face and Gesture Recognition, 2004, pp. 895–900.
10. H. Zhou and T. Huang, "Okapi-Chamfer matching for articulate object recognition," in Proc. International Conference on Computer Vision, 2005, pp. 1026–1033.
11. Y. Wu, J. Lin, and T. S. Huang, "Analyzing and capturing articulated hand motion in image sequences," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, pp. 1910–1922, 2005.
12. Haibo Li, P. Roivainen, R. Forchheimer, 3-D motion estimation in model-based facial image coding, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 15, Issue 6, June, 1993.

Interdisciplinarity

My application is interdisciplinary

☐

An interdisciplinary research project is defined in this call for proposals as a project that can not be completed without knowledge, methods, terminology, data and researchers from more than one of the Swedish Research Councils subject areas; Medicine and health, Natural and engineering sciences, Humanities and social sciences and Educational sciences. If your research project is interdisciplinary according to this definition, you indicate and explain this here.

Click here for more information

Scientific report

Scientific report/Account for scientific activities of previous project

---

Budget and research resources

Project staff

Describe the staff that will be working in the project and the salary that is applied for in the project budget. Enter the full amount, not in thousands SEK.

Participating researchers that accept an invitation to participate in the application will be displayed automatically under Dedicated time for this project. Note that it will take a few minutes before the information is updated, and that it might be necessary for the project leader to close and reopen the form.

Dedicated time for this project

	Role in the project	Name	Percent of full time
1	Applicant	Haibo Li	30
2	PhD Student		80

Salaries including social fees

	Role in the project	Name	Percent of salary	2016	2017	2018	2019	Total
1	Applicant	72,141	30	408,000	418,000	428,000	439,000	1,693,000
2	PhD Student	27,000	80	437,000	480,000	529,000	581,000	2,027,000
	Total			845,000	898,000	957,000	1,020,000	3,720,000

Other costs

Describe the other project costs for which you apply from the Swedish Research Council. Enter the full amount, not in thousands SEK.

Premises

	Type of premises	2016	2017	2018	2019	Total
1	office	102,000	108,000	116,000	123,000	449,000
	Total	102,000	108,000	116,000	123,000	449,000

Running Costs

	Running Cost	Description	2016	2017	2018	2019	Total
1	travels	conference	40,000	40,000	40,000	40,000	160,000
2	small equipment	mobile devices	20,000	0	0	20,000	40,000
	Total		60,000	40,000	40,000	60,000	200,000

Depreciation costs

Depreciation cost	Description	2016	2017	2018	2019

Total project cost

Below you can see a summary of the costs in your budget, which are the costs that you apply for from the Swedish Research Council. Indirect costs are entered separately into the table.

Under Other costs you can enter which costs, aside from the ones you apply for from the Swedish Research Council, that the project includes. Add the full amounts, not in thousands of SEK.

The subtotal plus indirect costs are the total per year that you apply for.

Total budget

Specified costs	2016	2017	2018	2019	Total, applied	Other costs	Total cost
Salaries including social fees	845,000	898,000	957,000	1,020,000	3,720,000		3,720,000
Running costs	60,000	40,000	40,000	60,000	200,000		200,000
Depreciation costs					0		0
Premises	102,000	108,000	116,000	123,000	449,000		449,000
Subtotal	1,007,000	1,046,000	1,113,000	1,203,000	4,369,000	0	4,369,000
Indirect costs	439,000	466,000	497,000	530,000	1,932,000		1,932,000
Total project cost	1,446,000	1,512,000	1,610,000	1,733,000	6,301,000	0	6,301,000

Explanation of the proposed budget

Briefly justify each proposed cost in the stated budget.

Explanation of the proposed budget*

LKP 53.20%
Löneökning 2.50%

Kontor 12.07%

Indirekta Kostnader(TB)
KTH 25.22%
Skola 26.71%

51.93%

Other funding

Describe your other project funding for the project period (applied for or granted) aside from that which you apply for from the Swedish Research Council. Write the whole sum, not thousands of SEK.

Other funding for this project

Funder	Applicant/project leader	Type of grant	Reg no or equiv.	2016	2017	2018	2019

CV and publications

CV

CV Haibo Li

1. Higher education degree(s) (year, subject area)

1997: Master's Degree in Electrical Engineering, Nanjing University, P. R. China.

2. Doctoral degree (year, discipline/subject area, dissertation title and supervisor)

1993: Ph.D in Information Theory from Linköpings Universitet, Dept. of Electrical Engineering (ISY), Sweden. Thesis title: Low Bitrate Image Sequence Coding (Supervisors: Prof. Ingemar Ingemarsson, Information Theory, and Prof. Robert Forchheimer, Image Coding)

3. Postdoctoral positions (year and placement) no

4. Docent (Senior Lecturer) level (year)

1997: Docent in Information Theory from Linköpings Universite

5. Present position, period of appointment, percentage of research in the position

2012 - , Full Professor in Media Technology, Program Director of Master of Science Program in Media technology, School of Computer Science and Communication, at KTH Royal Institute of Technology, 70% research, 30% education

6. Previous positions and periods of appointment (indicate the type of employment)

- *1999-2012, Full Professor in Signal Processing, Department of Applied Physics and Electronics (TFE), Umeå University*
- *1997 - 1998: Associate Professor, Division of Image Coding, ISY, Linköpings universitet.*
- *1993-1997: Assistant Professor in Division of Image Coding, ISY, Linköpings universitet.*
- *1990-1993: Research Assistant, Division of Information Theory, ISY, Linköpings universitet.*

7. Interruptions in research **n**o

8. PhD and Post-docs

PhD

- Astrid Lundmark, Linköping University, 2001 (as second supervisor)
- Jacob Ström, Linköping University, 2002 (as second supervisor)
- Jörgen Ahlberg, Linköping University, 2002 (as second supervisor)
- Lena Klasén, Linköping University, 2003 (as second supervisor)
- Apostolos Georgakis, Umeå University, 2004
- Zhengrong Yao, Umea University, 2005
- Jiong Sun, Umeå University, 2006
- Hungson Le, Umeå University, 2007
- Ulrik Söderström, Umeå University, 2008
- Shafiq Ur Rehman, Umeå University, 2009
- Johannes Karlsson, Umeå University, 2010
- Karin Fahlquist, Umeå University, 2013
- Farid Abedan, Umeå University, 2014
- Shahrouz Yousefi, KTH, 2014

Post-docs

- Dr. Guangming Lu, Nanjing University, 2006-2007
- Dr. Dengyin Zhang, Nanjing University, 2007-2008
- Dr. Alaa Halawani, Albert-Ludwigs University of Freiburg, Germany, 2008-2010.
- Dr. Fei Li, Nanjing University, 2007-2008
- Dr. Fredric Lindström, Göteberg University, 2007-2008
- Dr. Danny Chen, CSIRO ICT Centre, Australia, 2010-2011
- Dr. Zhihan Lv, Université Denis Diderot (Paris VII), 2012-2013

9. Other information of importance to the application

- Awarded "Best Nordic PhD Thesis" in the field of Image Analysis and Pattern Recognition 1993-1994
- Best Paper Award, IEEE International Conference on Machine Vision, 2007.
- ACM/IEEE Best Paper Award, 2010
- Brave New Ideas, ACM Multimedia 2010
- Best Case Study Award, ACM CHI 2013
- Nominated to Best Paper Award, The 12th International Conference on Mobile and Ubiquitous Multimedia (MUM) 2013
- Winner of KTH and EIT ICT Labs Idea Competition ICT 2013
- Guest Editor of Special Issue on Context-Aware Networking and Communications, I and II, IEEE Communication Magazine, 2014
- Apostolos Georgakis, PhD student of Prof. Haibo Li, won industripriset 2006, in SSAB (Svenska sällskapet för automatiserad bildanalys) 2006.
- Johannes Karlsson, PhD student of Prof. Haibo Li, one of two finalists to Chester Carlsons Forskningspris 2011.
- Prof. Haibo Li and his team won, First Prize, Winner of Academic Business Challenge, 2009.
- Wawo Technology AB, a spinoff company from digital media lab, First Prize, Venture Cup Nord, 2009
- Wawo Technology AB, semi-finalist to Innovate!100 list of top 100 tech startups in the world.
- Vimio, a spinoff company from digital media lab was listed in London AIM Stock Market in 2005 with a value of 26 million pounds.
- AMT AB, a spinoff company was listed as one of ten most potential IPTV companies, 2007 by ABI Research**,** a world-famous marketing research firm based in New York.
- Prof. Haibo Li gave an invited talk in Red Herring Spring 2007, Silicon Valley, (a top VC conference)
- Prof. Haibo Li gave an invited talk in Focus Innovation, 2008, Göteberg.

10. VR Funding + EU Funding

- "Tactile Perception of Facial Expressions", Vetenskapsrådet(**VR**), Project Research Grant, 2005-2007, Co-Applicant
- "Vibrotactile Rendering of Facial Expressions", Vetenskapsrådet(**VR**), Project Research Grant VR, 2008-2010, Co-Applicant
- "Is Wyner-Ziv coding a core technique enabling next generation face recognition technology for large-scale face image retrieval?", Vetenskapsrådet(**VR**), Project Research Grant, 2010, Principal Applicant
- "Distributed Wyner-Ziv Video Coding for SWE – Share What I Experience", Vetenskapsrådet(**VR**), Industrial PhD Fellowship, 2007-2010, Principal University Applicant
- "Green Video Sharing", Vetenskapsrådet(**VR**), Research Link, 2009-2011, Principal Applicant
- "Automatic Assessment of Pain in Infants by Computer Facial Expression Analysis", Vetenskapsrådet(**VR**), Research Link, 2005, Principal Applicant
- **More than 6 EU projects**

Publication list

Publication of Haibo Li (Google Scholar Ciatation)

Most Cited IEEE Transaction Papers

Haibo Li, P. Roivainen, R. Forchheimer, 3-D motion estimation in model-based facial image coding, **IEEE Trans. Pattern Analysis and Machine Intelligence**, Vol. 15, Issue 6, June, 1993. **Citation: 418**.

Haibo Li, A. Lundmark, R. Forchheimer, Image sequence coding at very low bit rates: a review, **IEEE Trans. Image Processing**, Vol. 3, Issue 5, 1994. **Citation: 246**.

Haibo Li, R. Forchheimer, Two-view facial movement estimation, **IEEE Trans. Circuits and Systems for Video Technology**, Vol. 4, Issue 3, 1994. **Citation: 44**.

A. Lundmark, Haibo Li, Hierarchical subsampling giving fractal regions, **IEEE Trans. Image Processing**, Vol. 10, Issue 1, 2001. **Citation: 37**.

Haibo Li, R. Forchheimer, A transformed block-based motion compensation technique, **IEEE Trans. Communications**, Vol. 43, Issue 234, 1995. **Citation: 18.**

Book Chapter (2009-)

Shahrouz Yousefi, Farid Kondori, Haibo Li, “3D Gesture Analysis Using a Large-Scale Gesture Database”, Book Chapter, **Advances in Visual Computing**, Lecture Notes in Computer Science, Volume 8887, pp 206-217, 2014

Ulrik Söderström, and Haibo Li. Side View Driven Facial Video Coding. Advanced Video Coding for Next-Generation Multimedia Services**. InTech, pp.139-154, 2013. Citation: 2.**

Ulrik Söderström and Haibo Li, Asymmetrical Principal Component Analysis Theory and Its Applications to Facial Video Coding, **Effective Video Coding for Multimedia Applications**, p. 95-110, 2011

Jean-Paul Kouma and Haibo Li, Large-Scale Face Image Retrieval: A Wyner-Ziv Coding Approach, in Book of **New Approaches to Characterization and Recognition of Faces**, Editor: Peter Corcoran, Publisher: InTech, Aug, 2011, ISBN 978-953-307-515-0.

Johannes Karlsson, Tim Wark, Keni Ren, Karin Fahlquist and Haibo Li, “Applications of Wireless Visual Sensor Networks - the Digital Zoo”, Book chapter in **Visual Information Processing in Wireless Sensor Networks**, 2010

Shafiq ur Réhman, Li Liu, "iFeeling: Vibrotactile Rendering of Human Emotions on Mobile Phones", In Springer in Mobile Multimedia Processing : Fundamentals, Methods, and Applications **(LNCS 5960: State-of-the-Art Surveys), pp. 1-20, 2010. Citation: 12.**

Ulrik Söderström and Haibo Li, HD Video Communication at Very Low Bitrates, **Mobile Multimedia Processing: Fundamentals, Methods, and Applications**, p. 295–314, 2009

Shafiq ur Réhman, Li Liu and Haibo Li, How to Use Manual Labelers in Evaluation of Lip Analysis Systems? **Visual Speech Recognition: Lip Segmentation and Mapping**, IGI Global LNCS 5960: State of the Art Survey, 239-258, 2009

Journal Papers (2007-)

Z. Yao and Haibo Li, Tracking a detected face with dynamic programming, **Image and Vision Computing**, Vol. 24, No.6, 2007, **Citation**: 23

H. Le and Haibo Li, Fused logarithmic transform for contrast enhancement, **Electronics Lettters**, 2008, **Citation**: 11

U. Söderström and Haibo Li, Asymmetrical principal component analysis for video coding, **Electronics Lettters**, 2008, **Citation**: 8

Shafiq ur Réhman, J. Sun, Li Liu and Haibo Li, Turn your mobile into the football: rendering live football game by vibration, **IEEE Trans. on Multimedia**, Vol. 10, Issue 6, 1022-1033, 2009, **Citation:** 11

F. Lindström, Keni Ren, Haibo Li, Comparison of two methods of voice activity detection in field studies, **Journal of Speech, Language, and Hearing Research**, Vol.52, Issue 6, 1658-1663 2009, **Citation: 10**

C. Schuldt, F. Lindström, Haibo Li，I. Claesson, Adaptive filter length selection for acoustic echo cancellation, **Signal Processing**, Vol. 89, Issue 6, 1185-1194, 2009, **Citation: 19**

Johannes Karlsson, Adi Anani, Haibo Li, Enabling real-time video services over ad-hoc networks opens the gates for e-learning in areas lacking infrastructure, **International Journal of Interactive Mobile Technologies**, Vol.2, p17- 23, 2009, **Citation: 1**

Ulrik Söderström and Haibo Li, Representation bound for human facial mimic with the aid of principal, **International Journal of Image and Graphics**, Vol.10, Issue 3, 343-363, 2010, **Citation: 8**

Alaa Halawani, Haibo Li, Adi Anani, Building eye contact in e-learning through head-eye coordination, **International Journal of Social Robotics**, Vol.3, Issue 1, p95-106, 2011**, Citation: 3**

Alaa Halawani, Shafiq ur Réhman, Haibo Li, Adi Anani, Active vision for controlling an electric wheelchair, **Intelligent Service Robotics**, Volume 5, Issue 2, April 2012, Pages 89-98. 2012. **Citation: 3**

Shahrouz Yousefi, Farid Abedan Kondori, Haibo Li, Camera-based Gesture Tracking For Real 3D Interaction Behind Mobile Devices, **The International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)**, Vol. 26, No. 8, 2012. **Citation: 2**

Farid Abedan Kondori, Shahrouz Yousefi, Haibo Li, Direct 3D head pose estimation from Kinect-type sensors. **Electronics Letters**, Vol. 50, No. 4, pp. 268-270. 2013, **Citation: 35**

Shahrouz Yousefi, Farid Abedan Kondori, Haibo Li, Experiencing Real 3D Gestural Interaction with Mobile Devices, **Pattern Recognition Letters**, Vol. 34, No. 8, 2013, pp. 912-921 , 2013, **Citation: 4**

Bo Li, Söderström, U., Ur Réhman, S., Haibo Li, Restricted Hysteresis Reduce Redundancy in Edge Detection. **Journal of Signal and Information Processing**, 4(3B): 158-163, 2013, **Citation: 1**

W. Shao, Q. Ge, H. Deng, Z. Wei and Haibo Li. Motion Deblurring using Non-stationary Image Modelling. **Journal of Mathematical Imaging and Vision**, Sept. 2014. **Citation: 0**

W. Shao, Q. Ge, Z. Gan, H. Deng and Haibo Li. A Generalized Robust Minimization Framework for Low-rank Matrix Recovery. **Mathematical Problems in Engineering**, Volume 2014. **Citation: 0**

Zhihan Lu, A. Halawani, Haibo Li, “Hand and Foot Gesture Interaction for Handheld Devices”, **ACM Transactions on Multimedia Computing Communications and Applications**. 10/2014; 11:1-19. **Citation: 0**

Shahrouz Yousefi, Haibo Li, 3D Hand Gesture Analysis through a Real-time Gesture Search Engine，**International Journal of Advanced Robotic Systems**，2015. **Citation: 0**

Qi Ge, C. Li, W. Shao and Haibo Li. A hybrid active contour model with structured feature for image segmentation. Vol. 108, **Signal Processing**, March 2015. **Citation: 1**

Conference Papers (2009-)

Shahrouz Yousefi, Farid Abedan Kondori, Haibo Li, 3D Gestural Interaction for Stereoscopic Visualization on Mobile Devices, In Proceeding of The 14th International Conference on Computer Analysis of Images and Patterns (**CAIP**), Seville, Spain, CAIP (2), Vol. 6855 Springer (2011) , p. 555-562, 29-31 August 2011, **Citation: 10**

Johannes Karlsson, Shafiq Ur Rehman and Haibo Li, Augmented Reality to Enhance Vistors Experience in a Digital Zoo, The 9th International Conference on Mobile and Ubiquitous Multimedia (**ACM MUM 2010**), 1-3 December, 2010, Limassol, Cyprus. **Citation: 8**

Guanming Lu , Xiaonan Li ; Haibo Li Facial expression recognition for neonatal pain assessment , International Conference on **Neural Networks and Signal Processing**, 7-11 June 2008, **Citation: 10**

Johannes Karlsson, Keni Ren and Haibo Li, "P2P video multicast for wireless mobile clients", International Conference on **MobiSys**, 2006. **Citation: 7**

Johannes Karlsson, J. Eriksson and Haibo Li, "Tracking and Identification of Animals for a Digital Zoo", The 1st IEEE/ACM **Internet of Things** Symposium, 18-20 December 2010, Hangzhou, China. **Citation: 7**

Fei Li, Min Zhou, Haibo Li, "A Novel Neural Network Optimized by Quantum Genetic Algorithm for Signal Detection in MIMO-OFDM Systems", IEEE Symposium on Computational Intelligence in Control and Automation (**SSCI 2011**- CICA) Proceedings, Paris France, pp.170-177, 2011, **Citation: 6**

Fei Li, Dongpo Zhu, Feng Tian, Haibo Li. "Improved Quantum Genetic Algorithm for Competitive Spectrum Sharing in Cognitive Radios", Lecture Notes in Electrical Engineering (**LNEE**) (ISSN: 1876-1100, Book Series), Springer-Verlag, pp.1009-1014, 2011, **Citation: 3**

Fei Li, L Zhou, Li Liu, Haibo Li. "A Quantum Search Based Signal Detection for MIMO-OFDM Systems", IEEE 18th International Conference on Telecommunications (**ICT 2011**) Proceedings, Ayia Napa Cyprus, pp.276-281, 2011, **Citation: 6**

Shafiq ur Réhman, Li Liu and Haibo Li, Lipless Tracking for Emotion Estimation, in Proc. of IEEE 3rd int'l Conf. on -Signal-Image Technology & Internet-based Systems (**SITIS' 2007**), pp.768-774, China, 2007, **Citation: 9**

Shafiq ur Réhman, Li Liu and Haibo Li , Manifold of Facial Expression for Tactile Perception, in Proc. of IEEE International Workshop on Multimedia Signal Processing (**MMSP07**), pp. 239-242, Greece, 2007. **Citation: 4**

Söderström, Ulrik and Li, Haibo, Anonymous video processing for live street view, IEEE International Conferences on **Internet of Things**, and Cyber, Physical and Social Computing, pp. 109-113, **Citation: 2**

Abedan Kondori, Li Liu, Shahrouz Yousefi, Haibo Li. Head operated electric wheelchair. In Proceeding of the IEEE Southwest Symposium on Image Analysis and Interpretation (**SSIAI2014**), April 6-8, 2014, San Diego, California, USA. 2014, **Citation: 1**

Shahrouz Youse, Farid Abedan Kondori, Haibo Li, Bare-hand Gesture Recognition and Tracking through the Large-scale Image Retrieval, in the 9th International Conference on Computer Vision Theory and Applications (**VISAPP**), January, 2014, **Citation: 1**

Z. Lv, S. Feng, S. Rehman and Haibo Li, Foot motion sensing: augmented game interface based on foot interaction for smartphone**, CHI'14** Exteneded Abstracts on Human Factors in Computing Systems, pages 293-296, 2014, **Citation**: 2

Shahrouz Youse and Haibo Li, 3D Interaction through a Real-time Gesture Search Engine, 12th Asian Conference on Computer Vision (**ACCV**), 2nd Workshop on User-Centred Computer Vision, 2014, **Citation: 2**

A. Darvish, Haibo Li, and U. Söderström, Super-resolution Facial Images from Single Input Images Based on Discrete Wavelet Transform, International Conference on **Pattern Recognition**, 2014. , **Citation:**

Bo Li, Aleksandar Jevtic, Ulrik Söderström, Haibo Li, Fast edge detection by center of mass, The 1st IEEE/IIAE International Conference on Intelligent Systems and Image Processing 2013 (**ICISIP2013**): 103-110, **Citation: 1**

Bo Li, Ulrik Söderström, Haibo Li, Independent thresholds on multi-scale gradient images , The 1st IEEE/IIAE International Conference on Intelligent Systems and Image Processing, 2013 (**ICISIP2013**): 124-131, 2013, **Citation:**

Farid Abedan Kondor, Shahrouz Yousefi, Li Liu, Haibo Li. Active human gesture capture for diagnosing and treating movement disorders. In Proceeding of The Swedish Symposium on Image Analysis (**SSBA2013**), Gothenburg, Sweden, March 2013, **Citation:**

Kallberg, D. ; Seleznjev, O. ; Leonenko, N. ; Haibo Li , Statistical Modeling for Image Matching in Large Image Databases, Internet of Things **(iThings/CPSCom)**, 2011 International Conference on Cyber, Physical and Social Computing , 2011, **Citation:**

Shahrouz Yousefi, Farid Abedan Kondori, Haibo Li, Stereoscopic Visualization of Monocular Images in Photo Collections, The 2011 IEEE International Conference on Wireless Communications and Signal Processing (**WCSP**), Nanjing, China, p 9-11 Nov. 2011. **Citation: 1**

Shahrouz Yousefi, Farid Abedan Kondori, Haibo Li, 3D Visualization of Single Images using Patch Level Depth, In Proceedings of the International Conference on Signal Processing and Multimedia Applications, **SIGMAP** 2011, Seville, Spain, 18-21 July, 2011. **Citation: 2**

Shahrouz Yousefi, Farid Abedan Kondori, Haibo Li, Robust correction of 3D geo-metadata in photo collections by forming a photo grid, The 2011 IEEE International Conference on Wireless Communications and Signal Processing (**WCSP)**, Nanjing, China, Nov. 2011. **Citation:**

Shahrouz Yousefi, Farid Abedan Kondori, Haibo Li, Tracking Fingers in 3D Space for Mobile Interaction, In Proceeding of The 20th International Conference on Pattern Recognition (ICPR), The Second International Workshop on Mobile Multimedia Processing (**WMMP**), Istanbul, Turkey, August 2010. **Citation:**

Farid Abedan Kondori, Shahrouz Yousefi, Haibo Li, Real 3D Interaction Behind Mobile Phones for Augmented Environments, In Proceeding of The 2011 IEEE International Conference on Multimedia and Expo **(ICME),** International Workshop on Hot Topics In Multimedia, Barcelona, Spain, 15th July 2011. **Citation: 3**

Farid Abedan Kondori, Shahrouz Yousefi, Haibo Li, Samuel Sonning, Sabina Sonning, *3D head pose estimation using the Kinect*, The 2011 IEEE International Conference on Wireless Communications and Signal Processing **(WCSP),** Nanjing, China, Nov. 2011. **Citation: 1**

Farid Abedan Kondori, Shahrouz Yousefi, *Smart Baggage In Aviation*, In Proceeding of The 2011 IEEE International Conference on Internet of Things iThings-11, International Workshop on Cyber-Physical Systems and Applications **(CPSA-11),** 2011. **Citation: 1**

Shahrouz Yousefi, Farid Abedan Kondori, Haibo Li, *3D Visualization of Monocular Images in Photo Collections*, In Proceeding of The Swedish Symposium on Image Analysis (**SSBA**), Linkoping, Sweden, February 2011. **Citation: 1**

Farid Abedan Kondori, Shahrouz Yousefi, Haibo Li, *Gesture Tracking for 3D Interaction in Augmented Environments*, In Proceeding of The Swedish Symposium on Image Analysis **(SSBA)**, Linkoping, Sweden, February 2011. **Citation: 1**

Karin Fahlquist, Johannes Karlsson, Keni Ren, Li Liu, Haibo Li, Shafiq Ur Rehman and Tim Wark, ” Human Animal Machine Interaction: Animal Behavior Awareness and Digital Experience”, **ACM Multimedia 2010** , 25-29 October 2010, Firenze, Italy. **Citation: 3**

Johannes Karlsson and Adi Anani, “Ad hoc Networks as a tool for E-Learning”, The International Workshop on New Achievements in the e-Learning’s Domain **(WAEL'09),** 16-18 June 2009, Umeå, Sweden. **Citation: 1**

Johannes Karlsson, Jean-Paul Kouma, Haibo Li, Tim Wark and Peter Corke. "Demonstration of Wyner-Ziv Video Compression in a Wireless Camera Sensor Network". The 9th Scandinavian Workshop on Wireless Ad-hoc & Sensor Networks (**ADHOC'09).** 4-5 May 2009, Uppsala, Sweden. **Citation: 1**

Johannes Karlsson, Jean-Paul Kouma, Haibo Li, Tim Wark and Peter Corke. "Poster Abstract: Distributed Video Coding for a Low-Bandwidth Wireless Camera Network". The 5th European conference on Wireless Sensor Networks (**EWSN 2008**). 30 January - 1 February 2008, Bologna, Italy. **Citation: 1**

Tim Wark, Peter Corke, Johannes Karlsson, Pavan Sikka, Haibo Li and Philip Valencia. "Real-time Image Streaming over a Low-Bandwidth Wireless Camera Network". The Third International Conference on Intelligent Sensors, Sensor Networks and Information Processing **(ISSNIP 2007).** 3-6 December 2007, Melbourne, Australia. **Citation: 1**

Johannes Karlsson, Tim Wark, Philip Valencia, Michael Ung, Hiabo Li and Peter Corke. "Demonstration of Image Compression in a Low-Bandwidth Wireless Camera Network". The 6th International Conference on Information Processing in Sensor Networks (IPSN 2007). 24-27 April 2007, Cambridge (MIT Campus), Massachusetts, USA. **Citation: 1**

Al-Naser, Mohammad, Haibo Li and Söderström, Ulrik, Reconstruction of occluded facial images using asymmetrical principal component analysis, 18th IEEE International Conference on Systems, Signals and Image Processing (**IWSSIP)**, pp. 276-279, **Citation: 1**

Söderström, Ulrik and Li, Haibo, Anonymous video processing for live street view, IEEE International Conferences on Internet of Things, and Cyber, Physical and Social Computing, pp. 109-113, **Citation: 2**

Källberg, D., Leonenko, N., Seleznjev, O., and Li, H. (2011a). Statistical modeling for image matching in large image databases. In: Proceed. IEEE International Conferences on Internet of Things, iThings/CPSCom, Dalian, China, 648-652. **Citation: 1**

Kouma, J.-P. and Li, H. Large-scale Face Images Retrieval: A distribution coding approach. International Conference on Ultra Modern Telecommunications. St. Petersburg, Russia. October 2009. **Citation:**

Kouma, J.-P. and Li, H. Large-scale Face Images Retrieval: A transform coding approach. 15th IEEE Symposium on Computers and Communications. Riccione, Italy. June 2010. **Citation:**

Kouma, J.-P. and Li, H. Large-scale Face Images Retrieval: A block coding approach. The 20th International Conference on Pattern Recognition. Istanbul, Turkey. August 2010. **Citation: 1**

Kouma, J.-P. and Li, H. Wyner-Ziv Coding for Large-Scale Face Image Retrieval. Swedish Symposium On **Image Analysis**. Halmstad, Sweden. March 2009. **Citation:**

Shafiq ur Réhman, Li Liu and Haibo Li, Vibrational Soccer: Tactile rendering of football game on mobile, in Proc. of IEEE International Conference and Exhibition on next Generation Mobile Application, Services and Technologies (NGMAST 2007), UK, 2007. **Citation: 1**

Shafiq ur Réhman, Li Liu and Haibo Li, Lip Localization and Performance Evaluation, in Proc. of IEEE int'l Conf. on Machine Vision (**ICMV' 2007),** pp.29-34, Pakistan, 2007. **Citation: 1**

Bo Li, Shafiq ur Réhman, Haibo Li, "i-Function of Electronic Cigarette: Building Social Network by Electronic Cigarette" , IEEE International Workshop on Cyber-Physical Systems and Applications (CPSA 2011), Dalian, China, 2011. **Citation:**

CV

Name: Haibo Li	**Doctorial degree:** 1993-11-15
Birthdate: 19651008	**Academic title:** Professor
Gender: Male	**Employer:** Kungliga Tekniska högskolan

Research education

Dissertation title (swe)
låg bitrate video kodning

Dissertation title (en)
Low Bitrate Image Sequence Coding

Organisation	Unit	Supervisor
Linköpings universitet, Sweden Sweden - Higher education Institutes	Institutionen för systemteknik (ISY)	Ingemar Ingemarsson

Subject doctors degree	ISSN/ISBN-number	Date doctoral exam
10209. Medieteknik	ISBN 91-7871-203-3	1993-11-15

Publications

Name: Haibo Li	**Doctorial degree:** 1993-11-15
Birthdate: 19651008	**Academic title:** Professor
Gender: Male	**Employer:** Kungliga Tekniska högskolan

Li, Haibo has not added any publications to the application.

Register

Terms and conditions

The application must be signed by the applicant as well as the authorised representative of the administrating organisation. The representative is normally the department head of the institution where the research is to be conducted, but may in some instances be e.g. the vice-chancellor. This is specified in the call for proposals.

The signature*from the applicant*confirms that:

- the information in the application is correct and according to the instructions form the Swedish Research Council
- any additional professional activities or commercial ties have been reported to the administrating organisation, and that no conflicts have arisen that would conflict with good research practice
- that the necessary permits and approvals are in place at the start of the project e.g. regarding ethical review.

The signature *from the administrating organisation* confirms that:

- the research, employment and equipment indicated will be accommodated in the institution during the time, and to the extent, described in the application
- the institution approves the cost-estimate in the application
- the research is conducted according to Swedish legislation.

The above-mentioned points must have been discussed between the parties before the representative of the administrating organisation approves and signs the application.

Project out lines are not signed by the administrating organisation. The administrating organisation only sign the application if the project outline is accepted for step two.

Applications with an organisation as applicant is automatically signed when the application is registered.