Parallel Distributed Face Search System for National and Border Security

Brian C. Lovell, Abbas Bigdeli, and Sandra Mau
NICTA and School of ITEE, The University of Queensland
Level 5, Axon Building, Staff House Rd, St Lucia
The University of Queensland, Australia
{Brian.Lovell,Abbas.Bigdeli,Sandra.Mau}@nicta.com.au

Abstract

The CCTV surveillance industry is undergoing a sea change due to the adoption of IP technologies. This is allowing the integration of a plethora of new cameras and other sensors into huge integrated networks. Adoption of IP technologies is presenting opportunities for scalable visual analytics that has the potential to add enormous value to entire camera networks. One such technology is scalable robust face search to identify persons of interest in large crowds. Not only are such systems required to work robustly in a wide variety of conditions, they must also be extremely fast and scalable to hundreds, if not thousands, of high definition camera nodes.

Developing and testing such technology is challenging and requires a combination of fast algorithms, distributed databases, mobile platform integration, parallel processing using distributed middleware such as ROS, and GPU acceleration using tools such as CUDA and OpenCL. In this paper we cover emerging system trends such as super-megapixel cameras, post incident digital PTZ, integration and fusion of video and non-video sensors, multimodal remote biometrics including face and iris on the move. Finally, recognition results will be presented from a formal face recognition trial in early 2011 in one of Asia's largest International airports.

1. Introduction and Background

As biometric technologies advance due to improved capture and search technologies, they are becoming widely adopted. Early adopters of biometric technologies include the national immigration and border protection services. For example, all foreigners entering the USA are fingerprinted and their faces are digitally photographed under controlled conditions. This procedure is completed manually by US border control officials at the arrivals desk and is very labour intensive. In Australia and New Zealand, passengers with biometric passports have the option of entering the country via the fully automated SmartGate[2] booths as shown in Figure 1. Smartgate uses facial verification technologies to match the identities of each passenger with their biometric chipped passport. However face verification is certainly not as mature and reliable a technology as the border agencies would like. Indeed, the face verification scanner gates at Manchester Airport were suspended in February 2011 after a man entered the UK using his wife's passport[11]. The couple had inadvertently switched passports, so the problem was detected when his wife tried to enter and was blocked because she had presented her husband's passport.

Figure 1. Australian Customs SmartGate System in Sydney Airport

1.1. The Need for Speed

Despite these legitimate operational concerns, automated gates are gaining in popularity not primarily because of enhanced security but because of improved passenger facilitation. Thus it is clear that successful new border technologies must be reasonably secure, but an overriding concern is that they must not impede the flow of passengers through the terminals.

A good example of a recently successful embedded border technology is the HBOX® Iris in Motion technology from Hoyos Corporation[6] as shown in Figure 2. The

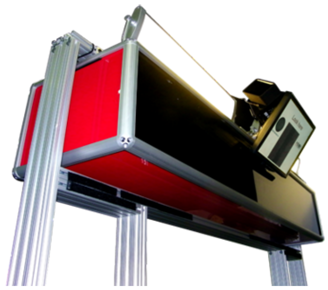


Figure 2. Hoyos Corporation's HBOX® Iris-in-Motion System

technology consisting of embedded processors, sensors and IR panels running multiple wave lengths. The HBOX® can recognise up to 50 persons per minute with close to 100% accuracy. A person is recognised by simply walking through the gate at normal speed whilst looking slightly upwards towards the high-speed infrared cameras. The success of the HBOX® can largely be attributed to the exceptional reliability of identification, high throughput, and ease of use.

On the other hand, HBOX uptake is hindered by the cost of the new equipment and the thorny problem of collecting yet another public biometric database. Reliability issues aside, it would be very convenient to use face recognition instead of, or in conjunction with, iris because 1) the databases already exist, 2) persons can be identified from a distance using the existing CCTV network, and 3) humans can verify matches without computer assistance.

The desperate need for embedded systems emerges from the huge data rates of modern high definition surveillance cameras as well as the numbers of cameras. Modern CCTV cameras are transitioning from standard definition (0.3 Mpixel per frame) to high definition (2.0 Mpixel per frame) which represents a 6-fold increase in data. While a simple single core implementation may cope with standard definition video, a much more sophisticated approach is required for high definition data rates.

Indeed one of the favoured CCTV solutions in research group is cameras of up to 16 Mpixels per frame. One advantage of these super megapixel cameras is that they cover 50 times the area of a standard definition camera at the same resolution. This gives the ability to surveil a large area such as an airport check-in and still capture high-quality face images of all passengers. Super Megapixel cameras can largely replace both PTZ and fixed cameras for indoor environments as they allow post-incident PTZ (up to 7x zoom) as well as camera sharing by multiple agencies without compromising inter-agency security. That is, multiple agencies (*e.g.* Customs and Police) can simultaneously look at different persons of interest without revealing who exactly is under observation.

Given 1) the streaming nature of video, 2) the large number of simultaneous video feeds in a major airport, and 3) the 24/7 nature of airport surveillance, practical technologies simply must operate in real-time. Finding and recognising faces in such enormously high definition video streams in real-time is a significant technical challenge.

Figure 3. Super Megapixel CCTV Camera (16 Mpixel)

1.2. Face Search Engine (FSE)

For several years the Advanced Surveillance Group in NICTA have been developing a face recognition technologies based on face matching using a technique called Multi-Region Histograms[15] (MRH). The advantages of this algorithm are 1) high speed, 2) small memory footprint, 3) ability to match low-resolution images as found in CCTV, and 4) insensitivity to pose, illumination, expression, focus, and misalignment. The computational advantages 1 and 2 mean that the algorithm is ideal for embedded systems implementations. Furthermore, performance advantages 3 and 4 mean that the algorithm is well-suited to both CCTV and mobile applications where image capture conditions are largely uncontrolled.

There has been a great deal of work on face recognition technologies over the years [19] including some on video-based recognition [9]. We have implemented and evaluated many of these state-of-the-art systems, but all methods we have tested to date fail to adequately address the pressing need for good recognition from uncontrolled low resolution image probes using uncontrolled low quality galleries. We suspect that the lack of research interest in low-resolution images is partly due to the emphasis in public face recognition benchmarking on both high-resolution images [13]. Such benchmarking yields impressive recognition rates that are simply not attainable in practical working systems. In our experience, face galleries collected by user agencies are much lower quality than face recognition benchmarking databases. Our system has been designed from the outset to cope with low quality mismatched probes and galleries.

We call the developed technology FSE (Face Search Engine). Due to its extremely high speed on large databases, the system resembles a Google-like search engine rather than a traditional recognition system — hence the name.

In the rest of the paper we discuss embedded system technologies to allow FSE to operate in real-time and on mobile platforms. Then we discuss the overall operation of the system. Finally, we describe results from a recent real world trial in an international airport.

2. Embedded Solutions

We use a number of embedded systems solutions to achieve the desired speed and parallelism in our face technologies. The following sections briefly describe techniques employed to achieve faster searches on large face databases. Note that we have a family of instantiations of the core technology that share the same FSE library code. In particular we have a Linux-based Forensic Face Search Engine with a browser user interface, a Windows-based real-time video face recognition system, a ROS[14] and Linux based multi-camera multi-core parallel distributed system, and several iPhone applications.

2.1. GPU Acceleration

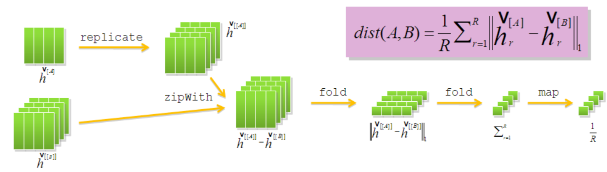
GPUs are massively multithreaded, many-core processors which are primarily used in computer graphics, video games and visual computing industry. Recently GPUs are being used in other application domains where exploiting fine-grained data parallelism can greatly accelerate the performance. Computer vision is one of those domains which can use GPU to significantly speedup performance due to their inherent data-parallel nature. Among different GPUs and Parallel processor manufacturers, nVIDIA was more successful, providing programmers with a programming model called CUDA (Compute Unified Device Architecture). Using CUDA, a GPU can be programmed as a co-processor which executes data-parallel kernels with a large number of threads. For development of our GPU Kernels, we used the OpenCL API for the CUDA architecture. OpenCL (Open Computing Language) is a framework to write GPU-accelerated programs which run across heterogeneous platforms.

Our face feature extraction engine is implemented on a GeForce GTX285 GPU. On videos with frame size of 640 by 480 pixels the performance of the GPU-based feature extraction engine is up to 40 times faster than the CPU-based implementation on an Intel Core 2 Duo.

In addition we have a GPU implementation of the core MRH algorithm using a new scalable parallel programming language currently being developed by collaborators in NICTA's Scalable Vision Machines Group[12]. The purpose of the language is to enable computationally intensive computer vision and machine learning algorithms to be written in a scalable manner. The code can then run on a cluster of highly-parallel heterogeneous platforms such as combined multi-core CPU multi-GPU platforms with FPGA acceleration. An example of the code is shown in Figure 4.

2.2. Distributed and Memory-Mapped Databases

A design goal of the Face Search Engine was the ability to near-instantaneously search millon person databases.




```
distances :: Array DIM2 -> Array DIM3 -> Array DIM1
distances histA histBs = dists
  where
    histAs = replicate (f, All, All) histA
    diffs  = zipWith (-) histAs histBs
    l1norm = fold (\a b -> abs a + abs b) 0 diffs
    regSum = fold (+) 0 l1norm
    dists  = map (/ r) regSum

    (f, r, h) = shape histBs
```

Figure 4. Scalable Programming with a Data Parallel language

The US Terror watchlist contains about 1 million names[16] already and is growing by about 20,000 per month. Even the state police in Queensland, Australia have a face database containing about 750,000 persons. Given the low computational load of the MRH algorithm and the simplicity of the match distance calculation, near instantaneous search returns are possible for database sizes of up to about 10,000 on a single CPU. The MRH face template is about 20k per person so a database of 5,000 persons is 100 Mbytes. Currently we memory map the face database, so search access is much faster than direct disk access or network database query. By replicating the database in the local memory of multiple cores, we provide a simple but manageable mechanism for cores to access the database without communication bottlenecks.

For larger datasets, we partition the search across several cores. Since the face search problem is simply a nearest neighbor search, it is trivially parallelizable. So the database is split into equal partitions and then each partition is assigned to a single core for searching. Once all cores have completed their partial searches, the partial rank ordered results are combined into a single result list. Smart indexing of the face search using, say, kd-trees may be possible at a later date, but so far we have not seriously investigated indexing approaches.

Moreover it is not clear whether FSE recognition performance on such large databases would be useful or practical for the user agencies. In practice, searches should be limited to most likely subsets of national databases based on gender, age, race, location *etc*.

2.3. Distributed Real Time Face Search Engine (FSE) System Using ROS

The Distributed Real Time FSE System is a framework for performing face searches across a network of cameras using a parallel distributed search engine utilises the Robot Operating System[14] from Willow Garage[17]. Originally developed at Stanford AI Laboratory under the name Switchyard, ROS provides hardware abstraction, device drivers, libraries, visualizers, message-passing, pack-

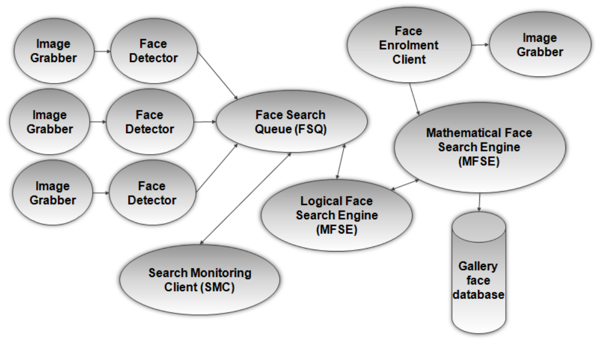


Figure 5. Architecture of the Face Search Engine

age management, *etc*. in an open source framework. ROS is a distributed middleware platform which allows tasks to be split among an arbitrarily large number of heterogenous computers.

The ROS-based FSE system (Figure 5) is able to extract faces from each camera in the network of cameras, and tag these faces with identities from a watch list using the FSE face recognition system. Each detected face is also logged for later matching against an individual of interest. Furthermore the framework supports a mode of operation where it can match faces detected in one camera against faces detected in other cameras on the network in real time, thereby allowing the operator to track the movement of a person through the network as it occurs.

The nodes of the system are split up into two groups, the tracking group and the recognition group. The tracking group is responsible for handling a single camera in the system. It extracts faces from its camera, and keeps track of the identity associated with each of those faces. The tracking group makes requests for face searches to the recognition group. This is a single system wide group of nodes responsible for extracting features from the faces sent to it, and comparing them against a watch list. The Recognition group also logs all faces it encounters for later analysis.

2.4. Smart Phone Implementation

Faster processors in smart phones have also enabled image processing on the move. The core FSE library as well as the OpenCV library have been cross-compiled for the Arm processor to provide face search on an iPhone embedded database derived from the phone's own contact list images (Figure 6). This approach can also be used for security on patrol using a suspect watchlist rather than a personal contacts list. Furthermore, smart phones can be used as part of a wider distributed network for real-time face detection and tracking by doing the face detection and feature extraction on the phone itself before searching against a database across a wider network.

In regards to performance, on the iPhone 4's underclocked processor at 800MHz, to process a video frame sized 270×480 pixels, OpenCV's face detection with

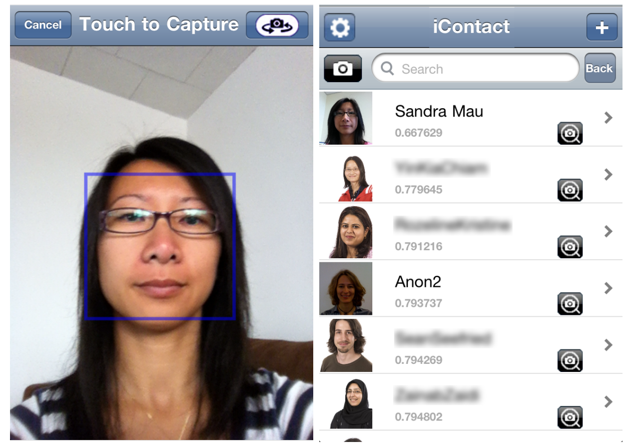


Figure 6. Sample application using Face Search Engine on a smart phone

haarcascade_frontalface_alt.xml requires just over 300ms to scan for a faceless frame (about half that with a face when using the option of finding only the biggest object). FSE requires around 4 seconds to extract the face histogram. This speed is typically good enough for social applications such as contact list search. With the next generation smart phones and handheld devices announced to have duo-cores[1] and OpenCL enabled GPUs[4], real-time image processing and distributed video analytics is shortly within reach.

2.5. Continuous Integration Environment

Coding large distributed parallel systems that run across multiple platforms and operating systems is a significant challenge for a small university research group. We currently use a Subversion server for revision control of the code. If the repository is changed, CMake[3] uses gcc[5] to compile all versions of our code for each platform. Finally a Hudson[7] continuous integration server runs a suite of unit tests on all platform builds to detect runtime errors. If any coding or runtime error is detected, an email instantly alerts the developer to the problem.

3. Distributed FSE System

The distributed FSE system performs fully automatic face enrolment by detecting and tracking each face in the field of view. Each face is tracked for several frames until tracking fails due to loss of face detection due to factors such as extreme pose or obscuration. These multiple frame tracklets are enrolled as individuals in the database so it is quite possible that the one person may have several tracklets. The importance of the tracklets is to ensure that we have multiple images corresponding to the same person.

A problem arises because many images are worthless for face recognition due to motion blur, face detection error, and obscuration of features. Addressing this, we have writ-

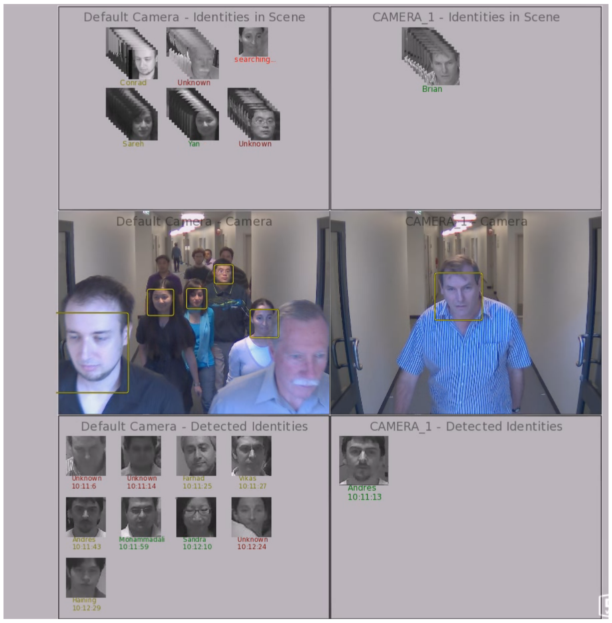


Figure 7. ROS Distributed FSE System Interface

ten a face quality module called libQ [18] which determines the quality of each face in the tracklet and selects the best of these for recognition purposes. Experiments on CCTV video data in [18] show that we obtain large improvements in recognition accuracy by processing just a small subset of high quality face images. As an added benefit, fewer images reduces the computational load. The end result is fully automatic real-time enrolment of a moving crowd of people passing by a CCTV camera. These individuals can then be matched against a watchlist of persons of interest, or matched against the tracklets from another camera. The former case gives us face identification whereas the latter case provides face recognition across a camera network. Figure 7 shows the system in operation. The image stacks represent the individual tracklets which are being matched to either a watchlist or tracklet list, or both lists simultaneously.

4. Undocumented Traveller Problem

A security technology is of no use unless it addresses a real and expressed user agency need. We have decided to initially target the undocumented traveller problem as it is well-suited to our technology as it works today. A major advantage of this particular face recognition problem is that persons are typically enrolled and then recognised within hours by cameras that we can specify to a large extent. Thus we don't need to deal with images from a huge variety of sources and resolutions (*c.f.* forensic matching) nor with photographs captured over significant periods of time (*c.f.* watchlist recognition).

This immigration problem occurs at most airports around the world and is extremely costly to taxpayers — thus there is a good financial incentive to address the problem. It arises when a person purchases a high quality false passport from a people smuggling organisation or equivalent. The person boards an aircraft on the false document and travels to a country where they wish to claim asylum.

After exiting the aircraft, the passenger destroys the false travel documents by, say, flushing them down a toilet where there is no surveillance. Later they present themselves at the immigration desk claiming a different identity (and often nationality) and request asylum. The cost of processing each asylum seeker is estimated at about USD80,000 which is borne by the taxpayer. Other countries are naturally reluctant to accept undocumented persons, so the country of arrival is forced to deal with the problem and expenses. Due to the *fait accompli* nature of arriving at the border without documents, the likelihood of success in seeking asylum seeking is quite high. This presents a clear danger to society due to the increased risk of accepting undesirable persons. The technical challenge is to track the person back from the immigration desk to the aircraft of arrival, so that the country of origin and travelling identity can be determined rapidly.

5. International Airport Trial

In January 2011 we had the opportunity to formally trial our real-time face recognition systems in a major Asian international gateway airport. The trials were funded by an Australian commercial partner in association with two local system integrators. Six systems integrators were represented. Other participants used well-known commercial face recognition technology. The trial was conducted formally and all participants were given the same DVD comprising surveillance video of airport passengers moving along a corridor past a 2 Mpixel camera. As this was a major international hub, the ethnicity and gender of the passengers was diverse and typical for this site. All processing for the trial including video enrolment was required to take place in a tight 120 minute window. Note that the video sequence alone lasted 72 minutes — so crowd enrolment was required to be both fully automatic and real-time.

Overall we enrolled about 4000 passenger tracklets from 7 large aircraft. Some groups of passengers were recorded multiple times on the DVD, which caused some confusion during the trial as noted below. Among these passengers were 15 employees who were our persons of interest (POI). The user acceptance criterion for the trial was 80% recognition (*i.e.*, 12 out of 15) where recognition was defined as the POI being found in the top 20 matches.

5.1. Results

During the official 120 minute trial period, we found 6 out of the 15 POIs — of these 4 POIs were recognised at rank 1 and the others were at ranks 6 and 8. However within an hour of the trial period, a recognition rate of 11 out of 15 was attained with simple system tuning. This unofficial result was recorded by 3 witnesses.

Complications include:

1. System crashing due lack of memory because of the sheer numbers being enrolled (solved by by pausing input video stream occasionally)

2. Due to these crashes, enrolment needed to be restarted where it left off in the video stream. Unfortunately there was confusion due to the multiple copies of certain groups of passengers, and this led to a section of the DVD being left completely unprocessed during the trial period. When this section was processed, two more POIs were found immediately.

3. Sometimes video quality was poor because of missing frames due to codec incompatibility during recording of the DVD. This problem caused passengers to apparently jump forward at times leading to person tracking errors and false enrolment.

4. Our tuning of the face recognition module was incorrect. We needed to use the low-resolution module instead of the high resolution module for best performance.

5. One POI was affected by frame skipping, so a good face image could not be obtained.

6. Two POIs did not show their faces at all — one man had a hat and had his head tilted down at all times and the other woman always had her hair obscuring her face.

In summary, due to the obscured faces of 3 POIs, the best a perfect face recognition system could have obtained in the trial would have been 12 or possibly 13 out of 15, so our actual recorded rate of 11 out of 15 was a satisfying outcome.

6. IFSEC Award

On Monday the 16th of May 2011, the Face in the Crowd system developed by the NICTA Advanced Surveillance Group won the prestigious IFSEC major category of CCTV System of the Year (excluding Cameras and Lenses) [8]. IFSEC is the world's largest global annual security event, uniting over 25,000 security professionals with more than 600 world leading companies. The IFSEC awards are simply the most prestigious prizes in the security industry. The

Figure 8. IFSEC2011 Award for CCTV System of the Year (excluding Cameras and Lenses) via commercialisation partner iOmniscient Pty Ltd.

Face in the Crowd system was entered into the competition by commercialisation partner iOmniscient Pty Ltd. Note that the system that actually won the award was an earlier multi-threaded windows-based software implementation rather than the newer and more powerful, parallel-distributed Linux version described in this paper.

7. Conclusions

In this paper we have described the development of a scalable face search engine capable of scanning CCTV video for faces in real-time. To achieve real-time rates, we needed a gamut of embedded system technologies including Robot Operating System, GPU acceleration, and smart algorithm design. The face search achieved creditable results in international airport trials with an effective recognition rate of 11/12 or 91% when searching for POIs among several thousand passengers from 7 aircraft.

8. Future Work

In future work we intend to supplement face search with whole body search using clothing and appearance models

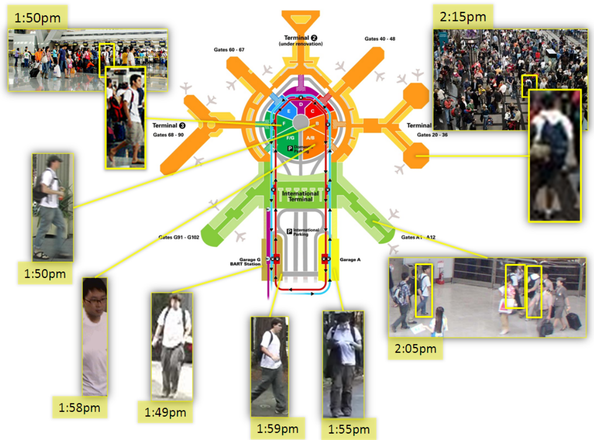


Figure 9. Proposed Face and Whole Body Search System

inside the ROS parallel distributed framework. The hope is to track and possibly identify all persons at all times within an airport or critical infrastructure environment as illustrated by Figure 9.

9. Acknowledgements

We would like to acknowledge the contribution of team members, Conrad Sanderson, Farhad Dadgostar, Mehrtash Harandi, Shaokang Chen, Ting Shan, Terence Smith, Jayden Platell, and Ian Cullinan. This project is supported by a grant from the Australian Government Department of the Prime Minister and Cabinet. NICTA is funded by the Australian Government's Backing Australia's Ability initiative, in part through the Australian Research Council.

References

[1] Apple Computer. iPad Features. http://www.apple.com/ipad/features/. [Last visited: 4-Mar-2011]. 120

[2] Australian Customs. Smartgate: How it works. http://www.customs.gov.au/site/page5831.asp. [Last visited: 4-Mar-2011]. 117

[3] CMake. Cross Platform Make. http://www.cmake.org. [Last visited: 4-Mar-2011]. 120

[4] CUDA Chess. OpenCL Enabled GPU. http://blog.cudachess.org/2011/01/ipad-2-with-opencl-enabled-gpu-retina-display/. [Last visited: 4-Mar-2011]. 120

[5] GNU. GNU Compiler Collection. http://gcc.gnu.org. [Last visited: 4-Mar-2011]. 120

[6] Hoyos Corp. Hoyos Corporation HBOX. http://www.globalrainmakersinc.com/Products/BusinessProducts/HBOX.aspx. [Last visited: 4-Mar-2011]. 117

[7] Hudson. Hudson Continuous Integration Server. http://hudson-ci.org. [Last visited: 4-Mar-2011]. 120

[8] IFSEC. IFSEC2011 Award Winners. http://www.ifsec.co.uk/Content/Finalists/9/. [Last visited: 30-May-2011]. 122

[9] V. Krüger, S. K. Zhou, and R. Chellappa. Integrating video information over time. example: Face recognition from video. In *Cognitive Vision Systems, LNCS*, volume 3948/2006, pages 127–144, 2006. 118

[10] B. C. Lovell, A. Bigdeli, and S. Mau. Embedded face and biometric technologies for national and border security. In *Seventh IEEE Workshop on Embedded Computer Vision Workshop at CVPR2011*, pages 117–122, 2011. Invited Paper.

[11] R. Massey. Face recognition scanners at Manchester Airport suspended after man re-enters Britain on his Wife's passport. http://www.dailymail.co.uk/news/article-1357949/Manchester-Airport-security-blunder-Woman-boards-flight-strangers-boarding-card.html, Feb 2011. [Last visited: 4-Mar-2011]. 117

[12] NICTA. NICTA Scalable Vision Machines. http://www.nicta.com.au/research/research_themes/embedded_systems/scalable_vision_machines. [Last visited: 4-Mar-2011]. 119

[13] P. J. Phillips, W. T. Scruggs, A. J. OToole, P. J. Flynn, K. Bowyer, C. L. Schott, and M. Sharpe. Frvt 2006 and ice 2006 large-scale experimental results. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(5):831–846, 2010. 118

[14] ROS.ORG. Ros documentation. http://www.ros.org/wiki/. [Last visited: 4-Mar-2011]. 119

[15] C. Sanderson and B. C. Lovell. Multi-region probabilistic histograms for robust and scalable identity inference. In *ICB, LNCS*, volume 5558, pages 199–208, 2009. 118

[16] R. Singel. U.S. Terror Watch List Surpasses 900,000 Names, ACLU Estimates UPDATED. http://www.wired.com/threatlevel/2008/02/us-terror-watch/. [Last visited: 4-Mar-2011]. 119

[17] Willow Garage. Robot operating system. http://www.willowgarage.com/pages/software/ros-platform. [Last visited: 4-Mar-2011]. 119

[18] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell. Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition. In *CVPR Biometrics Workshop*, pages 81–88. IEEE, June 2011. 121

[19] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003. 118