ISSN 2043-0167

Selected extended abstracts from the 7th Annual Postgraduate Conference

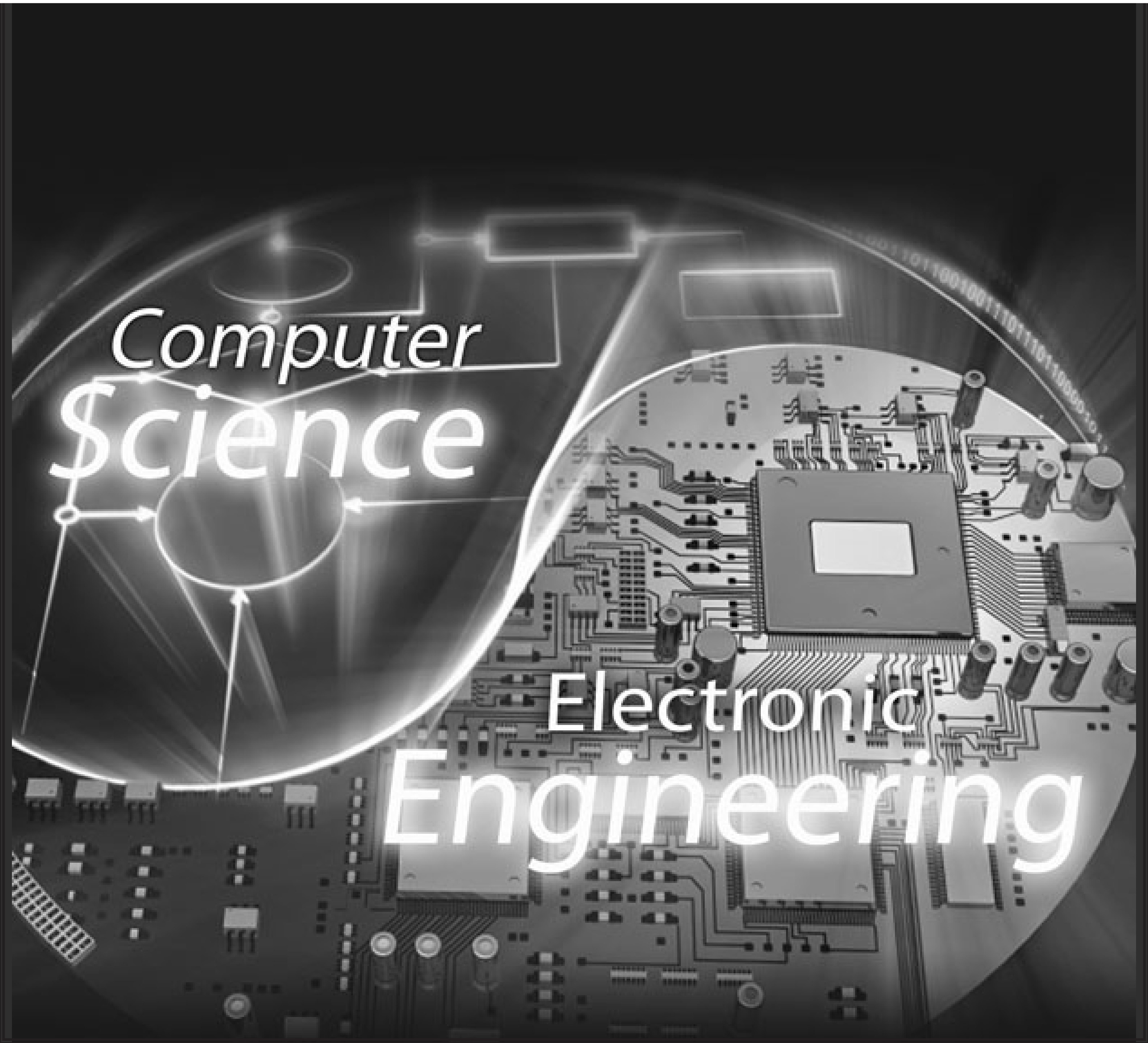
Editors: Stuart A. Battersby and Samuel Pachoud

Queen Mary
University of London

EECSRR-09-03

December 2009

School of Electronic Engineering
and Computer Science



Selected extended abstracts from the
7th Annual Postgraduate Conference

Department of Computer Science
Queen Mary University Of London

Contents

Clarification at a distance
Arash Eshghi

Generating customised experimental stimuli for visual search using Genetic Algorithms shows evidence for a continuum of search efficiency
Milan Verma

Recognising individuals in disjoint camera views
Bryan James Prosser

Lip feature selection based on similarity
Samuel Pachoud

Being spaced out
Stuart A. Battersby

Quantitative information flows
Jonathan Heusser

Auditory interaction with and through diagrams
Oussama Metatla

Social competence in schizophrenia
Mary Lavelle

Word reordering in statistical machine translation
Sirvan Yahyaei

Health & fitness project: A case study for application of Bayesian network in modelling time risks in projects
Vahid Khodakarami

Clarification at a distance

Arash Eshghi

Interaction, Media and Communication Research Group,
Department of Computer Science,
Queen Mary University of London,
arash@dcs.qmul.ac.uk

Multi-party conversations often lapse into sequences of dyadic interaction. During these sequences the other parties take on the status of Side Participants (SP) to the dyadic exchange [3]. Experimental evidence shows that despite their secondary role SP's can still have an impact on the form of the interaction between the primary participants (see e.g. [6]). In a recent corpus analysis of multi-party interactions we additionally showed that SP's can directly access the common ground developed between the primary participants using elliptical expressions; even though they have not, by definition, actively collaborated in constructing it. In fact, in this corpus study SP's were indistinguishable from primary participants in the forms of ellipsis they could use to access a prior dyadic exchange [2].

Despite this evidence, there is a strong intuition that SP's are different from the primary participants as to how they can access the common ground built up between the primary participants. “they [SP's] have to be satisfied with clearing up misunderstanding in natural breaks in their [primary participants'] talk” [1]. A SP often needs to wait for them to carry out their presentation and acceptance phases - and possibly resolve all the subquestions therein - before they can interject. Anecdotally, it seems that SP's frequently use techniques for re-raising context – avoiding highly elliptical expressions or in the case of anaphora and definite references, giving further descriptions of the discourse entities referred to – in order to access the prior context, which is at this point 'too far back' for elliptical access.

In this paper we explore the limits on this process through an experimental study of Clarification Ellipsis (CE, Elliptical clarification requests/repair). In particular we are concerned here with a special class of CE, namely Reprise Fragments (RF). These are questions that repeat part of a previous utterance in order to clarify it, such as below:

A: Did you speak to Mary yesterday?
B: Mary?

Using a chat tool technique described by [4] we are able to insert 'spoof' clarification ellipses with SPs as their apparent origin. This can be done without disrupting the dialogue and without being detected by any of the participants.

We use this technique to firstly to test the acceptability of RF's at different distances from the target turns containing the antecedent expression. (In dyadic dialogue RF's almost always immediately follow their antecedents [5]). In light of corpus evidence such as below we suspect that these fragments should be possible at higher distances (5 here) in multi-party dialogue:

C: What does cutest spelling mean? (1)
B: oh, she spelled cutest um with an I_, (2)
C: oh, okay. (3)
B: so that that's just something I pointed out. (4)
D: oh yeah. (5)
A: Cutest? [*Gazing at D. Direct Addresee is D here.*](6)
D: E_S_T_ (7)
A: Thank you.[laugh] (8)

We also explore the relationship between possibility/felicity of RF's relative to topic changes within the dyadic exchanges. We investigate whether 'new' questions/topics - which do not count as sub-questions of ones introduced before within the dyadic exchange - will make the ones introduced and resolved before, inaccessible as possible antecedents of RF. If so, this situation too necessitates context re-raising on the part of SP's.

According to [5] RF's can have two different readings/interpretations. These are the Clausal Clarifying and Constituent Clarifying readings. So, for the utterance "Mary?" in the example above, these readings are respectively: "Are you asking whether I spoke to Mary of all people?", and "Who is Mary". Intuitively the first of the two should not be possible at distance, since the hearer needs to resolve it by reference to the whole of the antecedent utterance which is 'too far back'. In the second reading however, what is asked about is contained wholly within the RF itself. So, using the experimental technique described above we can also test this intuition, by analysing the responses that RF's get at different distances. These responses reveal how the addressee of the RF has interpreted it.

Since clarification requests play a crucial role in the grounding process, the results of this experiment will need to be incorporated in any adequate account of grounding in multi-party conversations.

References

1. H.H. Clark and E.F. Schaefer. Dealing with overhearers. 1992.
2. Arash Eshghi and P. G. T. Healey. Collective states of understanding. 2006.
3. Erving Goffman. *Forms of Talk*. 1981.
4. P.G.T. Healey, M. Purver, J. King, J. Ginzburg, and G.J. Mills. Experimenting with clarification in dialogue. 2003.
5. M. Purver. The theory and use of clarification requests in dialogue.
6. Michael F. Schober and Herbert H. Clark. Understanding by addressees and overhearers. *Cognitive Psychology*, 1989.

Controlling what we see in visual search

Milan Verma
Vision Group, Interaction, Media & Communication Group
Department Of Computer Science
Queen Mary, University Of London
milan@dcs.qmul.ac.uk

Abstract. We present a new computational paradigm that generates on demand psychophysical stimuli with user-defined levels of salience for use in visual search tasks. Combining a Genetic Algorithm (GA) with a biologically motivated model for image saliency [2] we are able to 'breed' a range of textured elements with custom levels of saliency for use in psychophysical experiments. Experimental evidence for visual search continuum theories have to date primarily utilised stimuli that have been either produced manually or are limited in their diversity. To show the effectiveness of our artificial intelligence based approach to tailored experimental stimuli creation we present new psychophysical studies showing for the first time an explicit and predictable continuum of search efficiency exists in human visual search.

1 Introduction

In a visual search task a target item may be defined either by a unique distinguishing feature (feature search) or by a combination of features (conjunctive search). In a conjunctive search task, the distracter typically shares at least one feature with the target and as a result, the target item is less easy to spot against the background of distracters and this usually results in a serial search for the target. Conversely, in a feature search task the target-distracter disparity enables the target to 'pop-out' against the suppressed background [1][3]. In this case, the search slope of the task has a gradient of zero as additional distracters are added, inferring that additional distracter items have no effect on the search time. This is taken to support the viewpoint that rapid pre-attentive processes can operate in parallel across the entire visual field [4]. Despite the idea of a parallel/search dichotomy being dismissed in several publications (most notably [5]), there does not exist strong psychophysically evidence ruling it out once and for all.

In this paper we address this important research question using an improved computational model for visual saliency detection [2] and applying techniques from artificial intelligence to generate experimental stimuli. The methodology allows the generation of synthetic textons pairs, in which spatial pop-out occurs with pre-selected, user defined, levels of salience, so providing a new probe with which to examine the serial vs. parallel distinction. A Genetic Algorithm (GA) is used to optimise the search for texton pair images using the saliency model as the fitness function. Subsequent psychophysical tests verify the effectiveness of this process.

2 Synthesising stimuli

The experimental stimuli we seek to evolve here comprise of a series of Dawkins Biomorph textons. They have simple production rules yet still provide a rich diversity of instantiations. We develop these biomorphs on a 4×4 lattice of cells, where each 4×4 cell contains either target or distracter textons (see Figure 1 for examples).

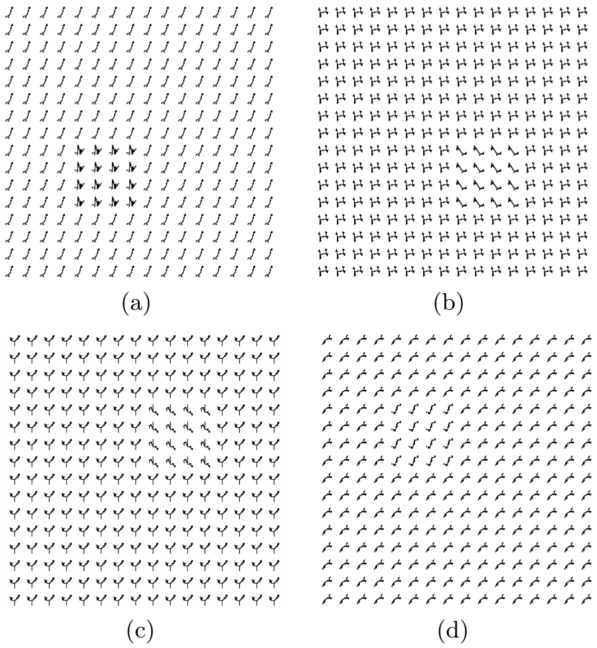


Fig. 1. Target-distracter images taken from the (a) 50th, (b) 33rd, (c) 15th and the (d) 6th generations of the GA. The images have a resolution of 400×400 pixels, with the target region having 4 possible locations around the centre of the image, (a) bottom-left, (b) bottom-right, (c) top-right and (d) top-left. Visual inspections show that as we move through the generations, from d-a, as the model determined saliency increases the pop out effect of the target region increases as expected.

Each 400×400 pixel image has two biomorph patterns, each having a unique 78-bit chromosome defining their appearance. Each biomorph is encoded with an initial 8-bits defining magnitude in 8 possible directions and the last 31-bits defining the directions chosen during each recursive step of the drawing algorithm. The initial Genetic Algorithm population is set at 12 chromosomes, elitism is set at 0.2 and the mutation rate at 0.04. Each subsequent generation evolves with the elite chromosomes along with randomly chosen chromosomes at random one-point cross-over positions. Figure 1 shows examples of the synthesised stimuli at various generations corresponding to changing levels of target-distracter saliency as measured by our saliency model.

3 Results

We psychophysically evaluated observer search performance on such stimuli, which had been generated to predefined target-distracter salience levels. Figure 2 presents the results of the experiment, showing the mean response time of predicted stimuli. The figure shows that response times have strong negative correlation to levels of salience. To test this interpretation, a Spearman's rho was conducted, which confirmed this as a main effect $[r = -0.72, n = 386, p<0.01$, with high levels of salience associated with lower response times.

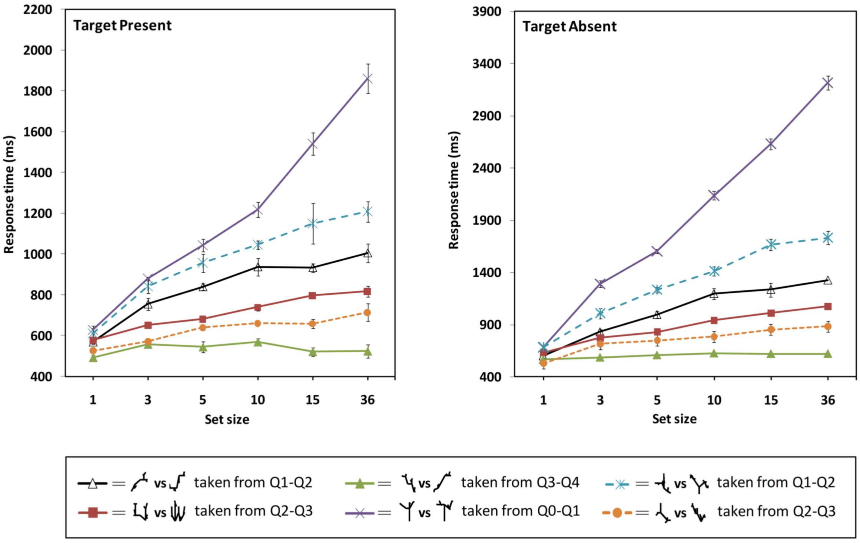


Fig. 2. Results of the search task using 6 different target-distracter pairings pulled from different sub ranges of the saliency continuum. The slopes indicate that as the saliency differential becomes less (i.e. texton pairs are selected from lower model saliency value ranges) that target detection time as a function of distracter set size increases. There also exists a continuum of slopes each slope reflecting the relative ranking of the target-distracter pairs value as selected from the saliency model range. Error bars indicate 1 s.e.m.

The methodology we have presented here allows the generation of target-distracter texton stimuli with pre set levels of saliency. The results of the experiment indicate that the methodology is psychophysically valid; our model can generate experimental stimuli to create a prescribed ranking of observer validated visual saliency. The results presented not only validate the mathematical model but provide useful evidence to inform the debate over the mechanisms supporting low level visual search. The response times from the experiment are wide ranging and using our predictable stimuli show a continuum of search efficiency where some target-distracter pairs are more efficient to search for than others. These results provide strong new evidence against a simple serial vs. parallel dichotomy in visual search. Furthermore, optimisation processes need a fitness function, and if the mathematical models used do indeed reflect the psychophysical reality, the stimuli obtained should recover the subject performance as predicted.

References

1. Beck, J.: Perceptual Grouping Produced by Line Figures. Perception and Psychophysics. In Perception and Psychophysics, **2**, (1967) 491–495
2. Itti, L., Niebur, E., & Koch, C.: A model of saliency-based fast visual attention for rapid scene analysis. Perception and Psychophysics. In IEEE Transactions on Pattern Analysis and Machine Intelligence, **20**,11, (1998) 1254–1259
3. Treisman, A. & Gelade, G.: A feature-integration theory of attention. In Cognitive Psychology, **12**, (1980) 97–136
4. Treisman, A., & Gormican, S.: Feature analysis in early vision: Evidence from search asymmetries. In Psychological Review, **95**, (1988) 15–48
5. Wolfe, J. M.: Visual search. H. Pashler (Ed.). In Attention, London: University College London Press, (1998) 15–48

Recognising Individuals in Disjoint Camera Views

Bryan Prosser

Computer Vision
Department of Computer Science,
Queen Mary, University Of London
bryan@dcs.qmul.ac.uk

1 Background:

Surveillance systems are becoming increasingly widespread in both public areas, such as streets and shopping centres, and private or protected areas, such as warehouses and military installations. As the number of cameras being used each year increases, the burden on CCTV operators increases, so much so that only a small number of cameras are being watched over in real time. This can result in many cameras being used only in reviewing an activity, such as a crime or shopping pattern, some time after the incident has actually occurred, rather than being used to prevent incidents or provide rapid response. One possible solution to this problem is to transfer some of the workload from the control room staff onto automated systems. A major task in such a move is to automate object recognition, particularly individual humans. The recognition of persons between camera views is known as Object Re-Identification. This is a process that humans are naturally good at and are able to quickly identify objects even in a crowded environments or when similar objects are present in the scene. Computers on the other hand do not have the wealth of prior knowledge that humans possess. Additionally to this, computer vision techniques are also much more sensitive to changes in colour, size and other recognition cues.

2 Recent Approaches:

Researchers have proposed a variety of methods for identifying individuals within a single camera view. Popular methods in the past have been based on facial recognition, gait and other appearance based models such as shape and colour. Each of these methods are limited individually by a reliance on a combination of set poses, size, scene illumination and object orientation. In a real world multi-camera system however, none of these factors are constant throughout the set of camera views. Of these approaches, colour is by far the most popular as it retains a certain amount of orientation and scale invariance, although it suffers greatly from illumination. Recently, Cheng et al[1] suggested that reducing the size of the colour space into major colours can be used across camera. The main problem with this approach is that reducing the feature space removes distance between similar objects, resulting in false identifications. Gheissari et al[2] suggest coupling shape and colour features. Their approach uses a decomposable triangular graph to segment similar-colour regions to provide spatial relationships between the colour regions. This approach attempts to reduces the effect of illumination and pose change between cameras, although the graph structure can only withstand a limited amount of pose change. Gilbert et al[4] also attempt to reduce the effect of illumination changes by producing a colour transfer function between camera views. They couple this with knowledge of the transition times between cameras to increase the chance of recognition, however their dynamic update approach leaves them a comparatively low success rate.

3 Our Approach:

Our approach is currently based on work from Javed et al[5] on Brightness Transfer Functions (BTF). Similar to [4], the BTF is designed to reduce the illumination differences between camera views by correcting colour observations for object re-identification. They assume that a certain percentage of the object in one camera will have brightness less than or equal to B_i and less than or equal to B_J in another camera view. From this they define the BTF between observations i and j as:

$$f_{ij}(B_i) = H_j^{-1}(H_i(B_i)), \tag{1}$$

where H_x is the cumulative histogram and H_x^{-1} is its inverse. A mapping function can then be produced to convert each colour value in each colour channel (RGB) from one camera view to another. Unlike [4], the subspace of all known BTFs is calculated via Principal Component Analysis. This subspace can then be used to compare new and existing objects by checking their BTF lies within the know subspace with Probabilistic Principal Component Analysis[6].

References

1. Cheng, E. D., Piccardi, M.:Matching of Objects Moving Across Disjoint Cameras IEEE International Conference on Image Processing (2006) 1769–1772.
2. Gheissari, N., Sebastian, T., Tu, P., Rittscher, J.: Person Reidentification Using Spatiotemporal Appearance. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (2006) 1528–1535.
3. Makris, D., Ellis, T., Black.: Bridging the gaps between cameras. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (2004) II-205–II-210.
4. Gilbert, A., Bowden, R.: Tracking objects across cameras by incrementally learning inter-camera colour calibration and patterns of activity. Proc. European Conference on Computer Vision. (2006) 125–136.
5. Javed, O., Shafique, K., Shah, M.: Appearance modelling for tracking in multiple non-overlapping cameras. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (2005) 26–33.
6. Tipping, M. E., Bishop, C. M.: Probabalistic principal component analysis. Journal of the Royal Statistical Society. (1999) 611-622

Lip feature selection based on similarity

Samuel Pachoud

Vision Group
Department Of Computer Science
Queen Mary, University Of London
spachoud@dcs.qmul.ac.uk

1 Introduction

Human perception is multi-sensory. In particular, one often uses two of our five senses: sight and hearing. Sight or vision describes the ability to detect electromagnetic waves within the visible range (light) by the eye and the brain to interpret an image as sight. Hearing or audition is the sense of sound perception and results from tiny hair fibres in the inner ear detecting the motion of a membrane. For perceiving facial emotion and behaviour, humans combine the acoustic waveform (audio information) and the movements of the lips, tongue and other facial muscles (visual information) generated by a speaker. The McGurk effect from [1] established this bimodal speech perception by showing that, when conflicting audio and visual stimuli is produced to an individual, the latter may assimilate a new stimulus, different from the two others. For example if you watch a talking head, in which repeated utterances of the syllable [ba] had been dubbed on to lip movements for [ga], normal adults reported hearing [da]. Effective human computer interaction (HCI) requires a multimodal speech recognition in order to make the computer interplays with human in the same way as human to human communication.

Such observations have motivated interest in developing systems for automatic recognition of visual speech. Research in this field aims to improve the speech recognition systems by taking advantages of the visual modalities of a speaker in addition to the usual audio modalities. The purpose of this combination is to improve the accuracy of a system compared to that depending on each single modality alone. One could consider three parts in constructing an audio-visual automatic speech recognition (AV-ASR) system: the front-end design, the audio-visual integration strategy, and the speech recognition method used. Figure 1 shows a general architecture commonly used for bimodal recognition.

Audio-visual automatic speech recognition systems introduced new and challenging tasks compared to traditional, audio-only ASR. The block-diagram of Figure 1 highlights these: in addition to the usual audio front-end (features extraction stage), visual features that are informative about speech must be extracted from videos of the speaker's face. Thus, in contrast to audio-only recognisers there are now two streams of features (one for each modality) available for recognition. The combination of the audio and visual streams should ensure

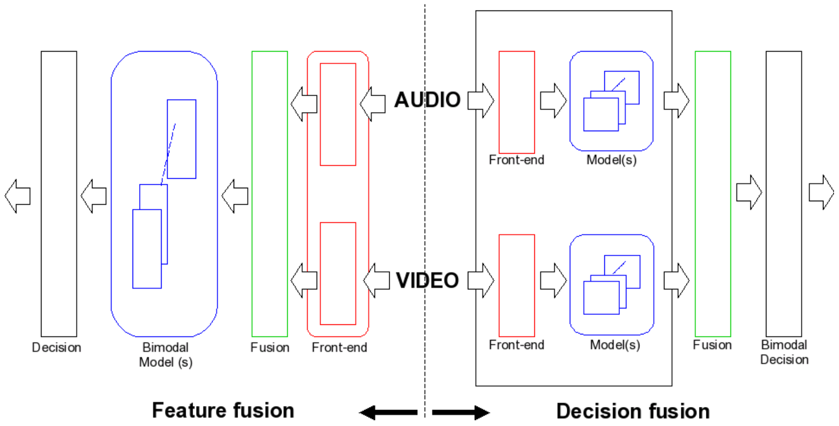


Fig. 1. General architecture for bimodal recognition. On the left, feature fusion class and on the right decision fusion class. With the first class, the same recognizer algorithms are used to concatenate audio and visual features. In the second one, separate recognisers are trained for the A+V features. This diagram shows the two additional challenging tasks to ASR system: the visual front-end design and the audio-visual integration.

that the resulting system performance is better than the best of the two single modality recognisers, and hopefully, significantly outperform it. Both issues, namely the visual front-end design and audio-visual fusion, constitute difficult problems, and they have generated significant research work among the scientific community. In this paper, we focus on addressing the problem of visual feature selection and extraction for lip-reading.

2 Related works

Visual speech recognition requires effective visual feature selection and extraction. The features need to be both robust and accurate.

To that end, there is a number of techniques that can be exploited, including Speeded Up Robust Features (SURF) [2], Scale Invariant Features Transform (SIFT) [3], Harris & Hessian Affine with SIFT descriptors (H&H) [4], Intensity Extrema Based Detector (IBR) and Edge Based Detector (EBR) [5], Maximally Stable External Regions (MSER) [6] and finally Salient Regions Detector (Salient Regions) [7].

Lowe's SIFT has been a popular method for feature extraction in object recognition. The SIFT descriptor computes a histogram of local oriented gradients around an interest point and stores the bins in a 128-dimensional vector (8 orientation bins for each of the 4×4 location bins). The features are considered to be invariant to image scale and rotation, and partially invariant (i.e. robust) to changing viewpoints and change in illumination. Several attempts have been made to further improve SIFT [8, 9].

SURF [2] is another interest point detector/descriptor scheme, which was inspired by the success of SIFT. At first, interest points are selected at distinctive locations in the image, such as corners, blobs and T-junctions. The most valuable property of an interest point detector is its repeatability, i.e. whether it reliably finds the same interest points under different viewing conditions. Then, the neighbourhood of every interest point is represented by a feature vector. This descriptor has to be distinctive and, at the same time, robust to noise, detection errors, and geometric and photometric deformations. Finally, the descriptor vectors are matched between different images.

A scale and an affine invariant interest point detector is developed by Mikolajczyk and Schmid in [4]. Their version of interest point detector combines the Harris corner detector [10] with automatic scale selection [11]. The Harris detector is based on the second moment matrix and its measure combines the trace and the determinant of the second moment matrix:

$$\textit{cornerness} = \det(\mu(\mathbf{x}, \sigma_I, \sigma_D)) - \alpha \mathrm{trace}^2(\mu(\mathbf{x}, \sigma_I, \sigma_D)) \tag{1}$$

where σ_I is the integration scale, σ_D is the differentiation scale. Local maxima of *cornerness* determine the location of interest points. The purpose of an automatic scale selection is to select the characteristic scale of a local structure, for which a given function attains an extremum over scales. Then they compute the operator responses for a set of scales for each point, and its scale selection operator in an image. They obtain an affine invariant image description which gives stable/repeatable results in the presence of arbitrary viewpoint changes.

Matas et al. [6] proposed a robust similarity measure for establishing tentative correspondences between image elements from two images with different viewpoints, using a robust wide baseline stereo from Maximally Stable Extremal Regions (MSER).

Tuytelaars and Van Gool [5] proposed a method to find a relatively sparse set of feature correspondences between wide baseline images. In each image, local image patches are extracted in an affine invariant way, such that they cover the same physical part of the scene (under the assumption of local planarity). The affine invariant way is either a geometric-based method (developed for curved and straight edges, EBR) or an intensity-based method (IBR). Then the patches or invariant regions are matched based on feature vectors of moment invariants that combine invariance under geometric and photometric changes.

Kadir et al. [7] described an extension to the region detector developed by Kadir and Brady [12]. The key principle underlying their approach is that salient image regions exhibit unpredictability, or 'surprise', in their local attributes and over spatial scale. The method consists of three steps: at first, calculation of Shannon entropy $\mathcal{H}_D(s)$ of local image attributes (e.g. intensity or colour) over a range of scales. Secondly, scales s_p are selected at which the entropy over scale function exhibits a peak. Finally, the magnitude change of the probability density function (PDF), $\mathcal{W}_D(2)$, is calculated as a function of scale at each peak. The final saliency is the product of $\mathcal{H}_D(s)$ and $\mathcal{W}_D(2)$ at each peak. The histogram of pixel values within a circular window of radius s is used as an estimate of the local PDF.

3 Proposed approach

We examine spatial features by finding similarities between image frames. To that end, we further exploit Boiman and Irani's approach [13], in which information-theoretic measures are derived for local and global similarities between a query (one or more signals) and a reference (one or more signals). For our purpose (find similarities between frames), the query or the reference signal are an image. Our approach works as follows: At first, we divide the query image in small macro-block (MB_q^i with ($i = 1, 2, \ldots, |\mathrm{MB}_q|$)) with partial overlapping. Then each MB_q^i is divided in a n blocks (B_q^j with ($i = 1, 2, \ldots, n$)). Those blocks have multi-scales (we currently use three different scales), so the number n depends of the scale. Then we look for similarities between each MB_q^i and the reference image. To perform this, we scan each B_q^j of each MB_q^i in the reference image. For each scale, the probability of a macro-block to be matched with the reference image is as follows:

$$P_s(\mathrm{MB}_q^i, ref) = \prod_i e^{-\frac{|\Delta d_i|^2}{2\sigma_d^2}} e^{-\frac{|\Delta l_i|^2}{2\sigma_l^2}} \quad (2)$$

where Δd_i and Δl_i are respectively the differences and local displacements between descriptors of B_q^j and their correspondents in the reference image. σ_d and σ_l are the only parameters we have to define. σ_d is determined empirically and σ_l is equal to the norm of the diagonal of a macro-block. Our descriptors are calculated using SIFT [3]. The total probability of a macro-block to be in the reference image, i.e. with the three different scales for B_q^j, is:

$$P_{tot}(\mathrm{MB}_q^i, ref) = \frac{1}{3}\left(\frac{(\sigma_{d1} + \sigma_{l1})}{2} P_{s1} + \frac{(\sigma_{d2} + \sigma_{l2})}{2} P_{s2} + \frac{(\sigma_{d3} + \sigma_{l3})}{2} P_{s3}\right) \quad (3)$$

Figure 2 shows a graph and an explanation of the process we have made until now. Picture (a) is the query image. Then we create small macro-blocks (MB_q). In (b), we see one of them in green and for the next pictures ((c) to (e)), the development is done with only one macro-block. Then, we extract each blocks (c) and we divide them in nine part as we see in (d) in blue. Next, we look for each MB_q separately in the reference image (e). We perform this without overlapping during the scanning of the MB_q as we show in yellow in (f). In (g), a bar graph of probabilities $P_{tot}(\mathrm{MB}_q^i, ref)$ (see Eqn. (3)) for each macro-blocks is drawn. The three highest probability of a MB_q to be in the reference image are showed in red in (h).

4 Experiments

4.1 Evaluation

The top-down methods (geometric-based) detailed in Section 2 are exposed above. With those methods the difficulty is to find points or features in a automatic way and to be able to match those points between different views or different frames. Consequently interest region detector algorithms were explored and

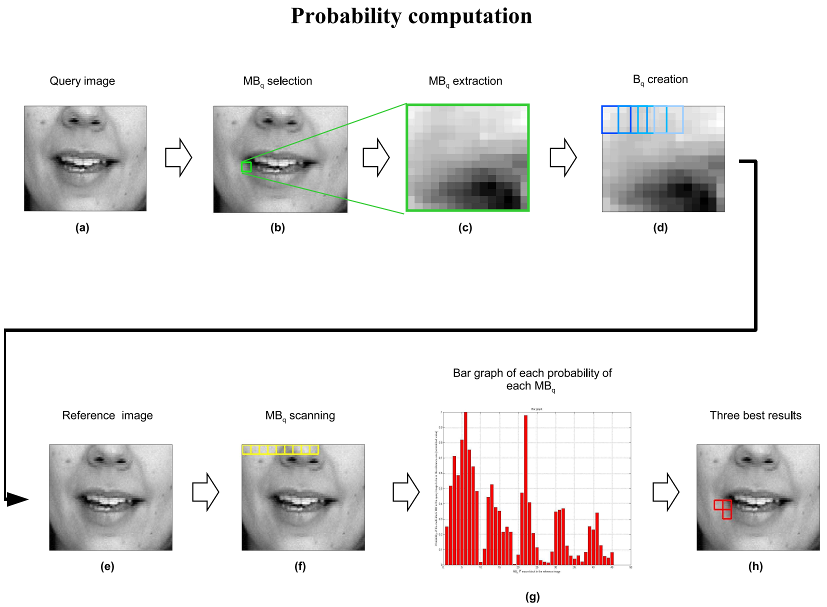


Fig. 2. Flow diagram of the process for MB_q modelling.

tested on the purpose of creating a robust visual feature extraction method. Detector and descriptor algorithms computed for local interest points were tested.

As only images are needed to test the interest region detector algorithms, only one or two frames per sequence of the In-house database [14] are sufficient to obtain the first results. Each frame are called by the name of the sequence (*dragana*, *zen*, etc...) and by the frame number: *sequence name_frame number*. Figure 3 to Figure 9 show results of whole interest region detector algorithms. For the display of each figure, the same construction is used: (a) is a female face from the file called *dragana*_0011 (it is the 11^{th} frame of the sequence *dragana*); (b) is the ROI of the latter (*dragana*_0011_*ROI*); (c) is a male face from the file called *zen*_0042; the ROI of *zen*_0042 is in (d) and named *zen*_0042_*ROI*; finally (e) is a picture of a room with multiple object (*room*).

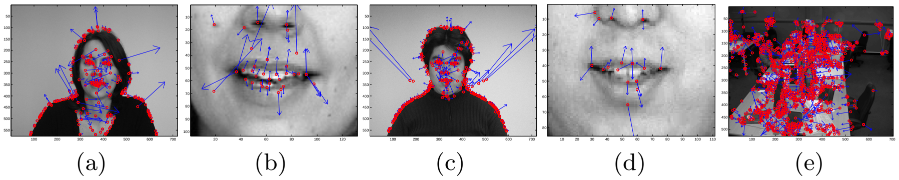


Fig. 3. Interest point detection with SIFT method [3]. A lot of interest point are detected even in the two ROI pictures.

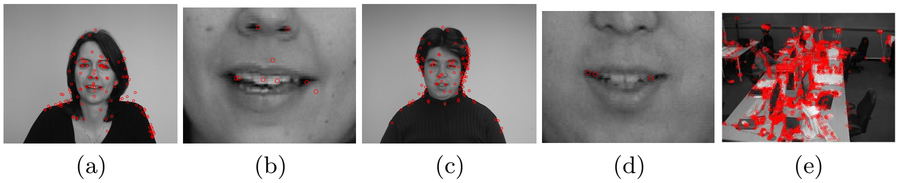


Fig. 4. Interest point detection with SURF method [2]. Few interest points are detected even in the two ROI pictures.

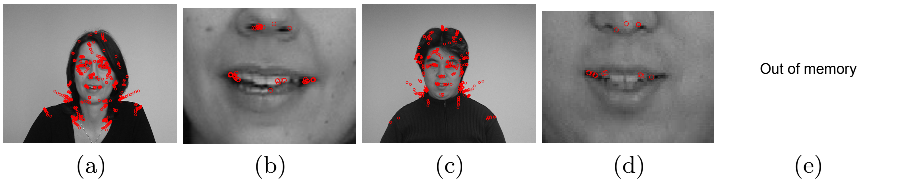


Fig. 5. Interest point detection with H&H method [9].

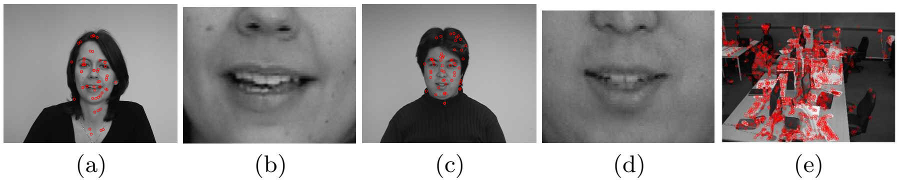


Fig. 6. Interest point detection with IBR method [5].

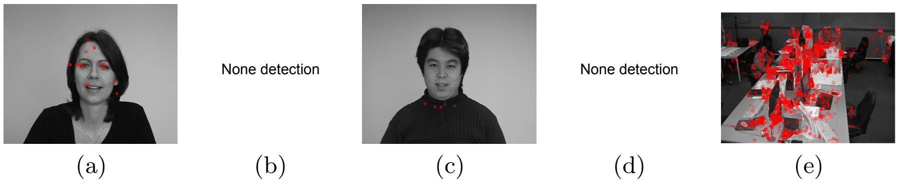
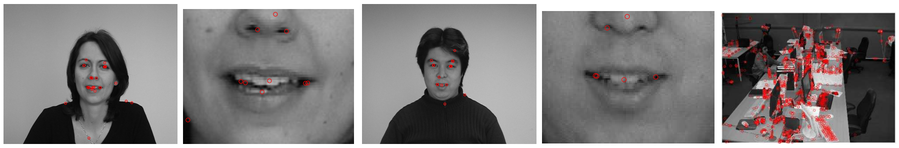


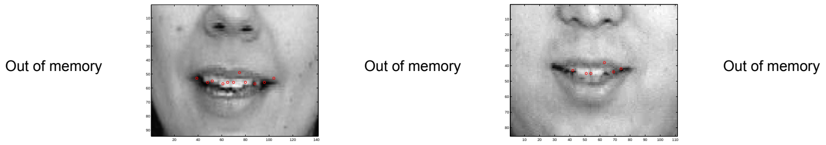
Fig. 7. Interest point detection with EBR method [5].

The experiments demonstrate that SURF method is not usable for lipreading: too few numbers of points are extracted with this method. With SIFT the results are slightly better but it is computationally expensive to compute the keypoints



(a) (b) (c) (d) (e)

Fig. 8. Interest point detection with MSER method [6].



(a) (b) (c) (d) (e)

Fig. 9. Interest point detection with salient regions method [7].

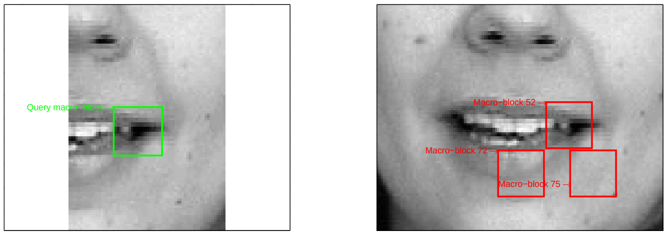
(the resulting descriptor is of dimension 128). Regarding the results, the other methods (MSER,EBR,IBR) are good for multiple object images but not enough accurate and efficient to be used for visual features for a lipreading system. The major problem with the automatic feature selection method is that there are not enough keypoints selected and moreover the accuracy is not optimal. Consequently the feature selection methods with interest point detector and descriptor are finally not adequate for lipreading purpose.

Then we tested our proposed approach on the same database. In Figure 10, the query image is compared to the reference one, resulting in the highest probability at the correct position for MB_q. In Figure 11 we change the scale keeping the same speaker: the query image contains the entire face (Figure 11 (a)) and the reference image (Figure 11 (b)) only the region of interest (ROI). Finally, in Figure 12, we have two different speakers, one male and one female for both the query frame and the reference frame respectively. We were able to find a part of the lip (MB_q^{71}) that is similar with MB_q (in green) shown in Figure. 12 (a).

It is evident from these three different experiments that our method is able to select and match regions between images and outperforms the other approaches for lipreading purposes. The regions contain the relevant information and the selection does not require any manual labelling. The model can also cope with query images from different speakers.

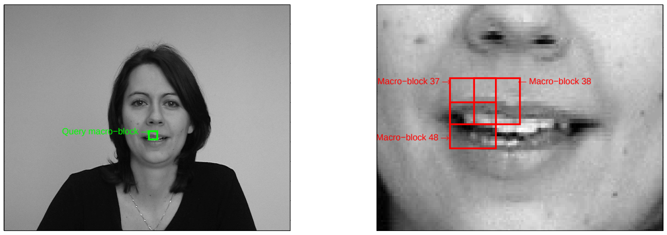
5 Conclusions

In this paper we present a novel approach for automatically selecting image features for lip-reading. Experimental results show that our method is able to detect regions that are similar between a query frame and a reference image under arbitrary scale change. Moreover, our model is also able to cope with query images from different speakers. Given both scale and speaker variation, the



(a) (b)

Fig. 10. Preliminary experimental result with the same frame for the query image and the reference one but translated. (a) is the query frame (MB_q in green). After the search in the reference frame, we see on (b) the three highest probability of a MB_q to be in the reference image with the following decreasing order: MB_q^{52}, then MB_q^{75} and MB_q^{72}

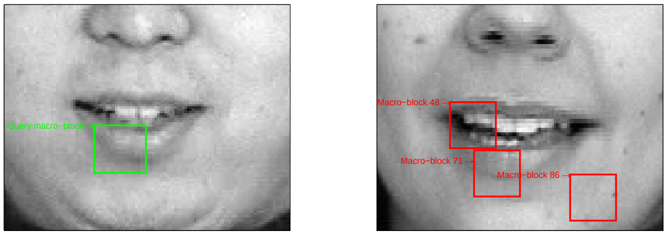


(a) (b)

Fig. 11. Sample results with the same speaker at different scale. (a) is the query frame (MB_q in green). After the search in the reference frame, we see on (b) the three highest probability of a MB_q to be in the reference image with the following decreasing order: MB_q^{38}, then MB_q^{48} and MB_q^{37}

expected region of lip movement is correctly found with the highest probabilities. It is important to point out that our model does not require any manual labelling of feature points or landmarks in training (i.e. unlike shape-based approach) and it is also not exhaustive in relying upon searching and matching dense texture features. Comparative evaluation shows that our method outperforms existing interest point based feature selection methods.

Lipreading is essentially the study of the movements of the lips, which involves a spatio-temporal space. The static spatial features approach presented in this paper can be extended to a representation in space and over time. This improvement is presented in [15]. Then in order to improve audio speech recognition system under noisy environment, a fusion between audio and visual features can be developed [16]



(a) (b)

Fig. 12. Sample results with two different speakers (male for the query image and female for the reference one). (a) is the query frame (MB_q in green). After the search in the reference frame, we see on (b) the three highest probability of a MB_q to be in the reference image with the following decreasing order: MB_q^{48}, then MB_q^{71} and MB_q^{86}

References

[1] H. McGruk and J. MacDonald: Hearing lips and seeing voices. Nature, **264** (1976) 746-748

[2] H. Bay and T. Tuytelaars and L. Van Gool: SURF: Speeded Up Robust Features ECCV (2006)

[3] D. G. Lowe: Distinctive Image Features from Scale-Invariant Keypoints IJCV, **60** (2004) 91-110

[4] K. Mikolajczyk and C. Schmid Scale & Affine Invariant Interest Point Detectors IJCV, **60** (2004) 63-86

[5] T. Tuytelaars and L. Van Gool Matching Widely Separated Views Based on Affine Invariant Regions IJCV, **59** (2004) 61-85

[6] J. Matas and O. Chum and M. Urban and T. Pajdla Robust Wide Baseline Stereo from Maximally Stable Extremal Regions BMVC (2002)

[7] T. Kadir and A. Zisserman and M. Brady An Affine Invariant Salient Region Detector ECCV (2004)

[8] Ke, Y. and Sukthankar, R. PCA-SIFT: a more distinctive representation for local image descriptors CVPR (2004)

[9] K. Mikolajczyk and C. Schmid A performance evaluation of local descriptors PAMI, **27** (2005) 1615-1630

[10] Harris, C. and Stephens, M. A Combined Corner and Edge Detection AVC (1988)

[11] T. Lindeberg Feature Detection with Automatic Scale Selection IJCV, **30** (1998) 77-116

[12] T. Kadir and M. Brady Saliency, Scale and Image Description IJCV, **45** (2001) 83-105

[13] O. Boiman and M. Irani Similarity by Composition NIPS (2006)

[14] P. Besson and V. Popovici and J-M Vesin and J-P Thiran and M. Kunt Extraction of audio features specific to speech using information theory and differential evolution EPFL - ITS (2005-018)

[15] S. Pachoud and S. Gong and A. Cavallaro Macro-cuboïd based probabilistic matching for lip-reading digits CVPR (2008)

[16] S. Pachoud and S. Gong and A. Cavallaro Video augmentation for improving audio speech recognition under noise BMVC (2008)

Being Spaced Out

Stuart A. Battersby

Interaction, Media & Communication
Department Of Computer Science
Queen Mary, University Of London
stuart@dcs.qmul.ac.uk

1 Introduction

Space is a highly salient concept in human interaction. It is influential at levels beyond our conscious understanding. Both where we are in space, and how we make use of it, are *integral* factors in our communication. We create spatial formations during interaction; these encapsulate jointly managed spaces that participants have mutually exclusive access to[6, 2, 3]. These spaces are then used for communicative activities, for example gesture.

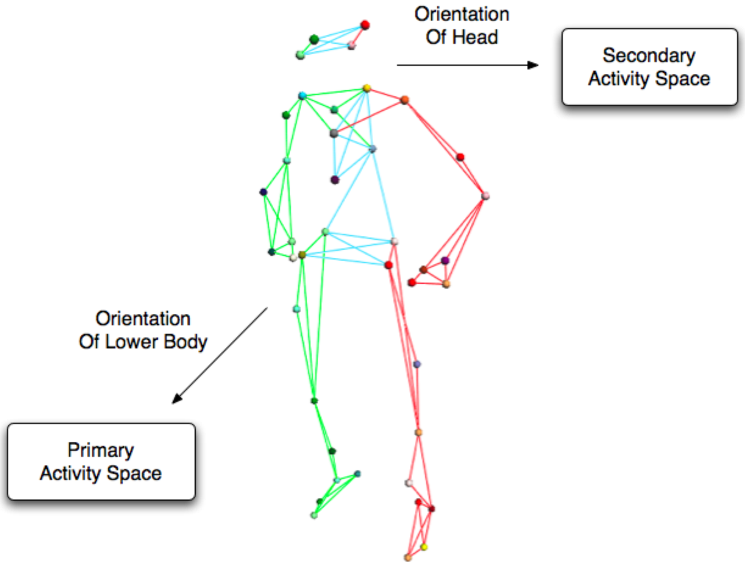


Fig. 1. A torqued body showing primary and secondary course of action spaces

This paper focuses on a spatial concept described by Emanuel A. Schegloff as *body torque* [7]. Body torque manifests itself in the body posture of participants during interaction. It is possible to orient the upper body and head differently to that of the lower body. This is know as a torqued body position. Schegloff suggests that these different orientations show differing levels of priority to different activities[1].

2 Example Data

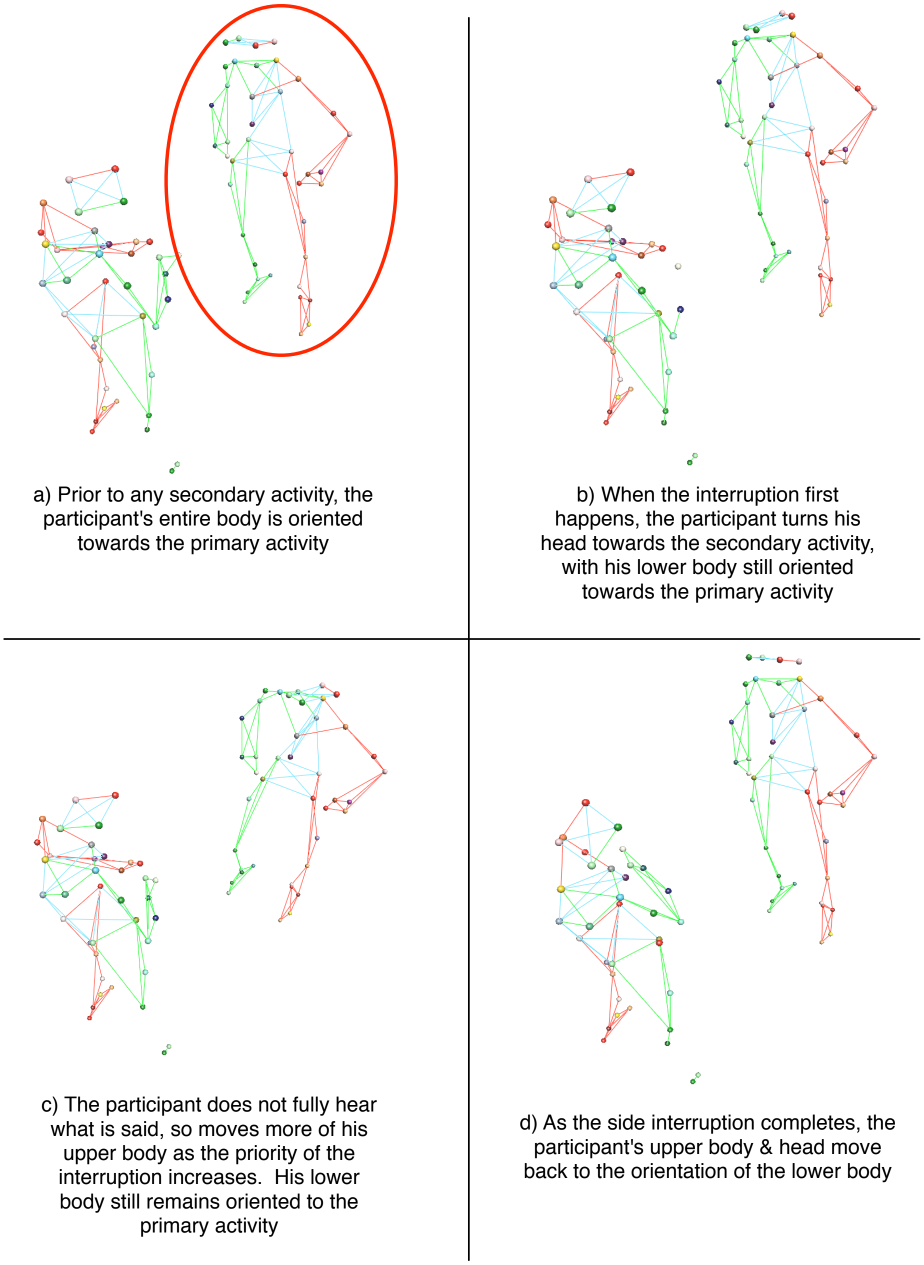
To highlight this concept with an example, we can examine three dimensional motion capture data from the Augmented Human Interaction laboratory at Queen Mary, University Of London (see Fig. 2 for the key stages of this data). This data gives us the precise three dimensional co-ordinates of body makers, placed on a participant to create a model of their body. This can then be represented visually using a wireframe figure. In a pilot study participants were asked to discuss the layout of their homes with each other; this was done to create a primary activity. Their interaction was then purposefully interrupted with the aim of creating a temporary secondary activity. This was done by the experimenter who entered the room to inform the participants that they had 2 minutes of discussion time remaining.

Prior to the interruption, the entire body of a participant was oriented forwards towards the jointly managed space and second participant (Fig. 2.a). This was the primary activity and had the participant's full focus. When the interruption occurred the participant's lower body remained fixed in its orientation, however the upper body (primarily the head) turned to orient to the secondary activity, the experimenter's interruption (Fig. 2.b). This put the body in a torqued position. During the interruption, the participant misheard what was said by the experimenter. When the information was repeated, the upper body was turned further towards the secondary activity showing an increase in priority(Fig. 2.c). Once the interruption was completed and the experimenter had left the room, the participant's body was released from torque and returned towards the original orientation to the primary activity(Fig. 2.d).

3 Theory of body torque

Schegloff states that the primary activity (in our example this was the discussion of the homes) is identified by the orientation of the lower body, with any secondary activity (in our example this was the interruption) being identified by the upper body's orientation. Given that we interact in a three dimensional spatial environment, these prioritised activities take place in separate spaces around the body (see Fig. 1).

[1] Schegloff's actual terminology was *courses of action.* This can be treated as synonymous with *activities*



a) Prior to any secondary activity, the participant's entire body is oriented towards the primary activity

b) When the interruption first happens, the participant turns his head towards the secondary activity, with his lower body still oriented towards the primary activity

c) The participant does not fully hear what is said, so moves more of his upper body as the priority of the interruption increases. His lower body still remains oriented to the primary activity

d) As the side interruption completes, the participant's upper body & head move back to the orientation of the lower body

Fig. 2. The key stages of the example data

These spaces are dynamic; the secondary activity may end, for example the end of the side interruption when the experimenter left the room. If this were to happen we would see the body released from its torqued position with the upper body falling back in line with the lower body and its activity space. However, if a shift of priorities were to occur, the interruption may become the primary activity, for example we could see the secondary activity become the primary; in this situation we would find again that the body would be released from its torqued position, however this time the lower body would reorient to fall in line with the upper body and its associated activity space. This would terminate any jointly managed space which had been formed at the lower body's previous orientation.

4 Conclusion

We have examined the theory of body torque proposed by Schegloff, and explored an example of motion capture data showing an occurrence of this. Whilst this is a subconscious activity, it is highly communicative and provides rich social cues which aid face to face interaction. Future work in this area will examine how this feature fits in with other theories of spatial interaction including f-formations ([3, 4]) and gestural topic spaces ([1, 5]). These studies will again make use of motion capture techniques, with a higher number of participants to draw statistically reliable results.

References

1. Emmorey, K.: Language, Cognition, and the Brain. Lawrence Erlbaum Associates. (2002)
2. Kendon, A.: The Role Of Visible Behaviour In The Organization Of Social Interaction. In: Social Communication and Movement. Academic Press (1973).
3. Kendon, A.: Conducting Interaction: Patterns Of Behavior In Focused Encounters. Cambridge University Press (1990).
4. Kendon, A.: The negotiation of context in face-to-face interaction. In Rethinking Context: Language as in interactive phenomenon. Cambridge University Press (1992).
5. Le Baron, C. and Streeck, J. Gestures, knowledge and the world. In Language and Gesture. Cambridge University Press. (2000).
6. Scheflen, A. E.: Human Territories: How We Behave In Space-Time. Prentice Hall (1976).
7. Schegloff, E. A.: Body Torque. Social Research. 65, 535–596 (1998).

Measuring Insecurity of Programs

Jonathan Heusser Pasquale Malacaria
Theory Group

Department of Computer Science
Queen Mary, University of London
`{jonathan,pm}@dcs.qmul.ac.uk`

Abstract. Recent advances in the theory and practice of quantitative information flow analysis allow, for the first time, to compute good approximations of security leaks of non-trivial programs. This paper presents a tool for a dynamic analysis of security leaks in programs. The tool first computes the precise leakage for a subset of the program inputs within a user-specified time. After exceeding this time limit, safe lower and upper bounds for all possible inputs are computed. These bounds also handle the case of non-terminating programs[1].

1 Introduction

Quantitative information flow analysis can provide a fine grained measure of a program's security or the lack of it (insecurity), i.e. the analysis computes results in numbers measuring security properties of a program. As opposed to that, qualitative analysis (e.g. non-interference) judge the quality of a program in comparison to predefined categories (e.g. secure/insecure). However, qualitative analysis is often too coarse in its decisions and would reject a lot of programs as insecure eventhough the probability of a leakage is very low.

Password protected access control is an often used example [7, 2, 3] to illustrate the idea of quantitative information flows. A password protected system *leaks data by its nature*, after all it tells you whether you entered the correct password or not. We take a 4-digit pin-protected system as example. The highest possible four digit pin number is 9999, thus the pin contains maximally $log_2(9999) \approx 13.3$ bits of information. When the attacker knows nothing about the pin then a single trial of entering a password reveals:

$$\frac{1}{2^{13.3}} log_2(2^{13.3}) + \frac{2^{13.3}-1}{2^{13.3}} log_2(\frac{2^{13.3}}{2^{13.3}-1}) = 0.00146$$

This is the weighted sum of the trials where the password was the correct one ($\frac{1}{2^{13.3}}$) and where it failed ($\frac{2^{13.3}-1}{2^{13.3}}$), which is an instance of Shannon's entropy

[1] This is a short version of the original paper which contains substantial case studies demonstrating the presented analysis, proof sketches and related work. Full version available upon request.

formula [9]. We can conclude that 0.00146 bits is very small compared to 13.3 bits of the secret and we *could* decide that this system is secure. Quantitative analysis allows that we detach the decision whether a program is secure or not from the actual security analysis; this is very important since different application areas have different security requirements.

In a qualitative setting, a password protected system like the one above would always get rejected as *insecure* since it leaks a very small amount of data. For real-world purposes this is not practical without using techniques which loosen the strictness of such an analysis.

1.1 Contribution

This paper presents a number of novelties in the research area of quantitative program analysis based on the information-theoretic semantics in [7]. We implemented the entropy-based formulas of [7] in an automated, dynamic program analysis. The precise information leakage of a while-program is calculated by evaluating the program on all or a subset of its inputs depending on the size of the secret. For all remaining, uncovered inputs safe lower and upper bounds are computed. These bounds also provide over-approximations of the leakage for the cases where the analysis does not terminate for a number of inputs.

2 Background

This section is giving a brief, informal overview of the theory presented in [7]. For a formal definition please consider the cited work.

The overall aim is to calculate the leakage of loops in a standard imperative programming language using Shannon's entropy formula[2] [9]. The initial assumptions for this calculation is that we have a program containing a while loop, a secret input containing the confidential data, a set of public variables, and an observer. The observer knows the program and the public variable assignments before and after the execution of the program. The task of the observer is to infer as much as possible of the secret input given the information he can gain by observing the program output. This gives rise to an informal definition of leakage

Definition 1 *The leakage of a program M is what an observer can infer of a secret input given the program code M and the full knowledge of the public variable assignment before and after the execution of the program.*

[2] This formula measures the average uncertainty of the output of a discrete random variable. For the probabilities $\{p_1, p_2, \cdots, p_n\}$ the entropy is calculated with $H(p_1, p_2, \cdots, p_n) = \sum_{i=1}^{n} p_i \log_2(\frac{1}{p_i})$

2.1 Random Variables and Programs

The language we are considering is a simple imperative language with assignments, conditionals, sequencing and a loop construct. Syntax and semantics of the language are standard and thus we omit them. For a command `M` of this language, we assume that there are two input variables `H` (confidential data), and `L` (public) which are equipped with a probability distribution. Thus, these variables can be considered as random variables, where the input to a program is the joint random variable $\langle \mathtt{H}, \mathtt{L} \rangle$. A deterministic program `M` can hence be seen as random variable itself: the output random variable where the probability of an output value of the program is the sum of the probabilities of all inputs evaluated by `M` to the value, i.e. $\mu(\mathtt{M} = o) = \Sigma\{\mu(h,l) | [\![\mathtt{M}]\!](h,l) = o)\}$

Example. Let us assume `M` is the assignment `l=h`, and that the variables hold 2 bit of information with uniform distribution (every value is equally likely). Now, the probability that `M` evaluates to a certain value e.g. 2 taken from the set of possible values $\{0, 1, 2, 3\}$ is $\frac{1}{4}$; every value for `h` is equally likely and the assignment random variable `M` is reflecting that input distribution.

For the random variable `M`, representing a program, we define an equivalence relation identifying all output states which are *observably equivalent.* This equivalence means that two states are counted as equivalent if all public variables hold the same value, e.g. their states after execution are not distinguishable. The relation is important to model how an observer keeps different output states apart.

Example. Let us assume `M(h,l)` is the conditional `if(h>l) l++ else l--` then `M(4,5)` and `M(4,3)` compute two equivalent output states, namely those where `l=4` after the execution of the command `M`.

2.2 Leakage of Loops

Let us consider a loop of the form `while e M`, where `e` is representing the Boolean guard and `M` is making up the random variables associated with the commands in the body of the loop. To the guard `e`, we associate *events* $e^{<i>}$ which denote the sequence of guard evaluations which are true for i times and always false from the $i + 1$th iteration on. The body `M` is just a random variable as described above, representing the commands in the body.

We denote $\mathtt{M}^i$ as the ith iteration of the body `M`. Now, we can describe the entropy of such a loop as the entropy of the probabilities of the events from `e` plus the entropy of $\mathtt{M}^i$ given the knowledge of $e^{<i>}$. Formally,

Proposition 1 *W(e, M) is the leakage of a loop* `while e M` *bounded by n iterations*[3]

$$W(e, M) = \underbrace{H(\mu(e^{<0>}), \cdots, \mu(e^{<n>}))}_{guard} + \underbrace{\sum_{1 \leq i \leq n} \mu(e^{<i>}) H(M^i | e^{<i>})}_{body}$$

[3] For simplicity this is the formula for loops without collisions [7]. We deal with collision in the analysis in section 3

Example. Consider the loop `l=0; while(l < h) l++` with 3 bit variables. There are only two events possible: the one where the loop exits, i.e. where `l >= h` and the one where `h` is already 0 and no iterations take place. This is represented by the events $e^{<n>}$ and $e^{<0>}$. Notice that the body can't leak anything because there are no confidential variables referenced in it. Thus, the leakage of the guard and therefore the whole loop is, assuming uniform input distribution of `h`, $H(\mu(e^{<0>}), \cdots, \mu(e^{<n>})) = H(\frac{1}{8}, \cdots, \frac{1}{8}) = \log_2(8) = 3$. As was expected, all 3 bits of the secret `h` leaked into `l` in this program.

3 Dynamic Analysis

The dynamic analysis formalizes the automation of Proposition 1, borrowing notation from relational algebra [1]. A projection $\pi_n(r)$ is defined as restricting tuple r to the n-th object of r ($\pi_n(R)$ is its extension to tuples); the selection $\sigma_\varphi(R)$ selects all tuples in R for which boolean formula φ holds. The denotational semantics of the program M is denoted as $[\![M]\!]$. It represents a state transformer (i.e. a map) which consists of a mapping from initial variable states to final variable states after the evaluation of the program. A state $\sigma \in \Sigma : \text{Var} \to \mathbb{N}$ is a mapping from variable identifiers to integer values[4]. The analysis chooses a subset of the possible inputs (following a user-chosen distribution) of the program M and stores the execution traces in the relation

$$R \subseteq \Sigma \times \mathbb{N} \times [\![M]\!](\Sigma)$$

given $r \in R$, $\pi_1(r)$ is an initial state for $[\![M]\!]$, $\pi_2(r)$ is the number of iteration it takes for M to terminate the loop on $\pi_1(r)$ and $\pi_3(r)$ is $\mathtt{obs}([\![M]\!](\pi_1(r)))$ the observable final state of M (i.e. the secret variables are not in $\pi_3(r)$). If $\pi_1(R) = \pi_1([\![M]\!])$ we note R by $[\![M]\!]_{\mathcal{R}}$.

Calculating an event $e^{<i>}$ from R is done by selecting a subset S_i of R, $S_i = \sigma_{n=i}(R)$, i.e. the subset of R with $\pi_2(R) = i$. Let the function $\text{pmf} : \Sigma \to \mathbb{R}$ be calculating the probability of a (input) state in Σ by a given probability mass function[5]. Using these definitions, the probability of $e^{<i>}$ is given by

$$\hat{\mu}(e^{<i>}) = \sum_{\sigma \in \pi_1(S_i)} \text{pmf}(\sigma)$$

The probability that a certain body iteration $M^i|e^{<i>}$ outputs a value v_i is given by the sum of probabilities of all initial states σ which evaluate to that value under M [7]. Formally, taking $v_j \in \pi_3(S_i)$ then the probability of v_j is

$$\mu_i(v_j) = \sum\{\text{pmf}(\sigma) \mid \sigma \in \pi_1(S_i) \wedge [\![M]\!](\sigma) = v_j\}$$

Having defined these quantities, R allows to compute the leakage formula from Proposition 1 as follows:

[4] Please do not confuse a state σ and the selection operation σ_φ

[5] In all case studies we assume that pmf is using uniform distribution

Proposition 2

$$W(e, M) = H(\hat{\mu}(e^{<0>}), \cdots, \hat{\mu}(e^{<n>})) + \sum_{1 \leq i \leq n} \hat{\mu}(e^{<i>}) H(\frac{\mu_i(v_1)}{V_i}, \frac{\mu_i(v_2)}{V_i}, \cdots, \frac{\mu_i(v_m)}{V_i})$$

where $\{v_1, \ldots, v_m\} = \pi_3(S_i)$ *and* $V_i = \sum_{j=1}^{m} \mu_i(v_j)$

To compute leakage for general loops we need to handle collisions. A collision is a state $v \in \pi_3(S_{i_1}) \cap \cdots \cap \pi_3(S_{i_k})$; denote by C the set of all such collisions in R. Then the leakage is given by

$$W_C(e, M) = W(e, M) - \sum_{v \in C} \mu(v) H(\frac{\mu_{i_1}(v)}{\mu(v)}, \ldots, \frac{\mu_{i_k}(v)}{\mu(v)})$$

where $\mu(v) = \sum_{i=1}^{n} \mu_i(v)$ and $\{i_1, \ldots, i_k\}$ is such that $v \in \pi_3(S_{i_1}) \cap \cdots \cap \pi_3(S_{i_k})$

To sum up, the tool computes the relation R by executing a program on (a subset of) inputs; then it computes the leakage $W_C(e, M)$. Proposition 2 is the computable version of Proposition 1.

3.1 Safe Bounds

Two independent problems often arise when analysing while loops:

- inputs are too large therefore we can't run the program on all inputs within reasonable time
- some inputs might make the loop non-terminating

Our bounds presented here provide a safe lower and upper bound if one (or both) of these problems occur. Both of the two cases lead to a reduction of the available events $e^{<i>}$. For the missing events we make safe worst case assumptions using maximum entropy [5] for the guard leakage and an estimation of what is possible to leak in the body of the loop.

Let R be a subset of $[\![M]\!]_{\mathcal{R}}$, which is complete i.e. $\forall i \in \pi_2(R)$. $\sigma_{n=i}(R) = \sigma_{n=i}([\![M]\!]_{\mathcal{R}})$. Suppose also $s = |\{i | i \in \pi_2(R)\}|$ and $t = |\{i | i \in \pi_2([\![M]\!]_{\mathcal{R}})\}|$. Then

Proposition 3 *Let* k *is the size of the secret and suppose that the leakage formula on* R *from Proposition 2 is*

$$L' = H(m_1, \ldots, m_s, q) + \sum_{i=1}^{s} m_i V_i$$

Then L' is a lower bound for the leakage of $[\![M]\!]_{\mathcal{R}}$. *An upper bound is*

$$\texttt{min}(k, L + q(k - L'))$$

where

$$L = H(p\ m_1, \dots, p\ m_s, \overbrace{\frac{q}{t-s}, \dots, \frac{q}{t-s}}^{t-s}) + \sum_{j=1}^{s} m_i V_i$$

with $p = \sum_{1 \leq i \leq s} m_i$ *and* $q = 1 - p$.
Moreover

1. *If the guard is not leaking the bound becomes* $\mathtt{min}(k, L' + q(k - L'))$
2. *If the body is not leaking the bound becomes* $\mathtt{min}(k, L)$
3. *These bounds are computable using the dynamical analysis.*

4 Conclusion

This paper presented an automated analysis of a theory quantifying information leakage in programs. The analysis consists of a dynamic and static part:

1. **Dynamic**. The program is run on a number of concrete inputs; the precise leakage for these runs is calculated by using the data of their execution traces.
2. **Static**. As a next step, lower and upper bounds safely approximate the maximal possible leakage for all remaining inputs using maximum entropy.

The analysis is based on concrete, not symbolic inputs.

Aknowledgments: We are very grateful to P. O'Hearn and H. Yang for comments on the full version of this work. Malacaria's research was supported by the EPSRC grant EP/C009967/1.

References

1. Alfred V. Aho and Jeffrey D. Ullman: Universality of data retrieval languages. POPL '79: Proceedings of the 6th ACM SIGACT-SIGPLAN symposium on Principles of programming languages.
2. David Clark, Sebastian Hunt, Pasquale Malacaria: Quantitative Analysis of the leakage of confidential data. Electronic Notes in Theoretical Computer Science 59, 2002
3. Michael R. Clarkson, Andrew C. Myers, Fred B. Schneider: Belief in Information Flow. CSFW '05: Proceedings of the 18th IEEE workshop on Computer Security Foundations, 2005:31-45
4. D. E. R. Denning: Cyptography and Data Security. Addison-Wesley, 1982.
5. E.T. Jaynes: Information Theory and Statistical Mechanics. Statistical Physics, 181, 1963.
6. Stephen McCamant and Michael D. Ernst: A Simulation-based Proof Technique for Dynamic Information Flow. PLAS 2007: ACM SIGPLAN Workshop on Programming Languages and Analysis for Security, 2007
7. Pasquale Malacaria: Assessing security threats of looping constructs. Proc. ACM Symposium on Principles of Programming Language, 2007.
8. Jonathan Millen: Covert channel capacity. Proc. 1987 IEEE Symposium on Research in Security and Privacy.
9. C. E. Shannon and W. Weaver: A Mathematical Theory of Communication. Urbana, IL: Univ. of Illinois press, 1963.

Auditory Interaction *With* and *Through* Diagrams

Oussama Metatla

Interaction Media & Communication group
Department of Computer Science
Queen Mary, University of London
oussama@dcs.qmul.ac.uk

Abstract. Human physical experience is enriched with sophisticated sensory capabilities. Its extension to the digital world as we currently know is still biased towards a single visual modality where much emphasis is put on graphical displays. Our research aims to explore alternative means for interacting with information, both in individual and collaborative usage scenarios, with a particular emphasis on using auditory display for accessing and manipulating diagrams. Investigating interaction techniques through alternative modalities will improve interactive experiences where vision cannot be relied on for optimum performance (mobile devices, multitasking, visual impairment). It will also increase our understanding of how to efficiently combine modalities to design sophisticated and usable interfaces.

Keywords: Human-Computer Interaction, Auditory Display, Diagrammatic Representations, Accessibility, Collaboration.

1 Introduction

Diagrams are a ubiquitous form of representation. As external representations [14], diagrams continue to be the subject of numerous investigations (e.g. [3, 5, 11, 14]), results from which are increasing our understanding of the properties that give this form of graphical representation such an integral role in supporting human cognition and communication. Most of such properties are strictly based on the visual characteristics of diagrams [3]. For instance, using the two dimensional plane to index information by location was shown to ease both searching and recognition and to decrease cognitive load associated with labeling [5].

The very graphical nature of such properties, however, means that the advantages of using diagrams are potentially lost when vision is not the optimum channel of communication to rely on. The tasks of constructing, exploring and retrieving information using diagrams provide a good context for investigating alternative interaction strategies. This is because designing effective means for presenting, navigating through and constructing such representations on digital devices is difficult without a visual display, yet important in contexts where the user's eyes are occupied or in the case of visual impairment. But how can a diagram be accessed through auditory means? And what is the optimal way to support auditory access to and manipulation of a diagram? Moreover, if we can efficiently access and manipulate a diagram through an audio-only interface, how is human communication and collaboration affected when it is supported through multimodal means?

These are the questions that form the core motivation to our investigations. We are mainly interested in exploring the design and evaluation of audio-only displays for supporting interaction with and through diagrams. We focus our investigations on relational diagrams, also referred to as nodes-and-links diagrams and typically used in Computer Science and Engineering disciplines [2]. This extended abstract briefly presents our previous, current and future studies pursuing this line of research.

2 Where do we come from? *- Previous work*

By interaction *with* a diagram we are referring to the process of inspecting existing diagrams, as well as constructing and editing new ones, for the purpose of discerning the structure of a given set of information, reasoning about a problem, or understanding a given procedure. In order to support auditory interaction *with* a diagram, it is therefore necessary to first devise a method to support *passive* interaction, whereby a user can inspect and navigate the information encoded in a diagram. Once this is established, we can then develop strategies for supporting *active* interaction with the diagram, whereby new information can be added onto it and edited out from it.

We thus first designed a model that maps the information represented in a relational diagram from its graphical form into a hierarchically organised structure. A multiple perspective hierarchical approach allows structured access to the information encoded in a relational diagram from a number of perspectives [8, 9] (see Fig. 1.). This model was based on Zhang's representational taxonomy of relational information displays (RIDs), which specifies the structures important for the systematic study of any RID [15]. Parts of the hierarchy are then displayed using a combination of continuous ambient sounds and momentary signal like speech and non-speech sounds [7, 8].

In a previous study [8], we evaluated the efficiency of the multiple perspective hierarchical approach in supporting passive interaction with relational diagrams. We explored how the hierarchy could be presented using two different auditory displays that vary in the amount of speech output each employs. The major outcomes of the evaluations were that:

- The multiple perspective hierarchical model allows for the relational information encoded in a diagram to be accessed and navigated using audio as the main means of information representation.

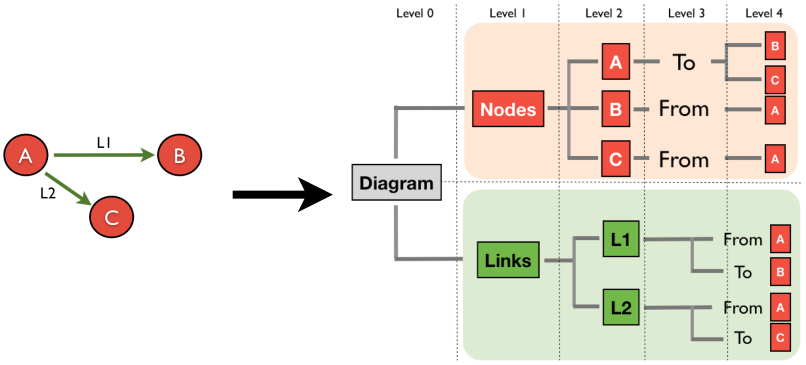


Fig. 1. Encoding a relational diagram as a multiple perspective hierarchy. Different parts of the hierarchy are displayed using speech and non-speech sounds. Shaded areas denote where continuous ambient sounds are audible.

- The substitution of verbal descriptions of parts of the diagram with nonverbal descriptions significantly improved performance times.

- The substitution of verbal descriptions of parts of the diagram with nonverbal descriptions did not compromise users' comprehension of those parts.

- The use of continuous ambient sounds provides contextual information that aid navigation.

Having extended on the concept of Interaction Traps [1] to systematically analyse users learning behaviour, we observed that users' learning rates were similar when participants interacted with a speech dominated and a non-speech dominated auditory interface. Learning rates were slightly hindered by orientation while navigating through the hierarchy. We thus further explored various auditory presentation strategies specifically aimed to improve the efficiency of user navigation and orientation within the hierarchical model. These consisted of varying speech parameters such as gender, speech rate and frequency, which were mapped to hierarchical depth. Results from these studies are not yet conclusive.

3 Why are we here? *- Current work*

After designing and evaluating a model that supports passive audio-only inspection of relational diagrams, we are currently exploring how to best support *active* interaction with this model. We have chosen Entity-Relationship (ER) diagrams as a context to explore strategies for constructing and editing a relational diagram in an audio-only interface [9, 10]. We designed two strategies for supporting such manipulations based on existing human-computer interaction techniques and inspired by Hutchins' analysis of interaction metaphors in interface design [4].

Much like sketching a diagram using pen and paper, the first strategy requires the user to locate the part of the diagram they wish to edit within the hierarchical model before executing the desired editing action on it (see Fig. 2.). We refer to this strategy as *Non-Guided* because the emphasis is put on the user as the main actor within a model-world where interactive expressions can be realised [4]. The hierarchy in this case represents both the model-world where actions are executed, as well as part of the language which expresses the editing actions.

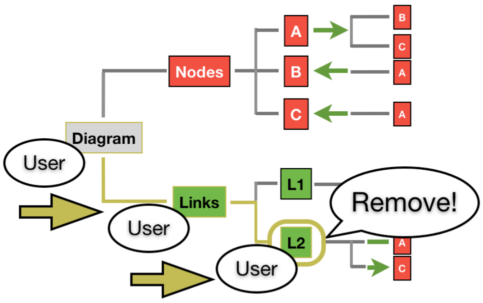


Fig. 2. In order to edit an item in the diagram in a Non-Guided interaction strategy the user must first locate it on the hierarchy before issuing the editing command.

Given that sound is being used as an alternative modality for manipulating an inherently graphical artefact, the representation, i.e. the diagram, and the means by which it is accessed, i.e. an auditory hierarchy, are essentially independent from one another. The second strategy exploits this independency to allow the user to edit any part of a diagram without having to 'physically' locate it on the hierarchy. (see Fig. 3.). The user can achieve this by engaging in a conversation-like interaction with the system, in which they express their desired to carry out an editing action, and then respond to a series of system prompts that guided through the necessary steps required to complete such action. We refer to strategy as *Guided* because it puts an emphasis on the system to act as an implied intermediary between the user and the world in which actions are taken [4].

Investigating the effect of these active interaction strategies on user's experience allows us to address a number of under-explored questions about the nature of interaction when manipulating a given representation through an alternative modality. The main questions that we empirically addressed in this part of the investigation were therefore: How to support active audio-only construction and editing of a diagram through the hierarchical model? And how effective is a given interaction strategy in supporting such activity? Detailed outcomes of this study can be found in [9, 10].

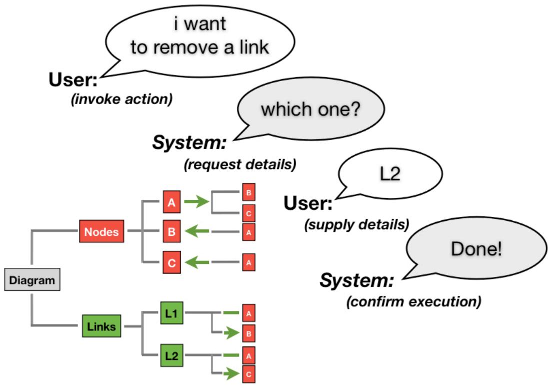


Fig. 3. In order to edit an item in a diagram using a *Guided* interaction strategy the user responds to a series of system prompts independently and away from the hierarchy.

4 Where are we going? *- Future Work*

By interacting *through* a diagram we are referring to the process of communicating one's thoughts and ideas to collaborating parties by means of constructing and acting on parts of or whole diagrams. These are common activities observed in group settings such as design teams or in a classroom [2].

Once mastered, a visual diagrammatic language can become an extremely efficient means for communication. There is on the other hand an evident lack of research on cross-modal collaboration in the fields of auditory display and computer supported cooperative work (mainly [6, 11, 12] to the best of our knowledge). The fact that very little work has been directed specifically towards cross-modal collaboration in this case forms both an advantage to the proposed research, in that it increases its originality, and a disadvantage, as no solid empirical and experimental background exists on which to base our work.

To further our investigations, we intend to use findings from our previous and current work on supporting auditory interaction with diagrams, to explore collaborative cross-modal problem-solving. If we can efficiently access and edit a diagram without looking at it using audio as the main means of interaction, then how are we to use it to collaborate with others who can see it and use visual means to construct and edit it? And how does the mutlimodal nature of such interaction affect both the process and outcome of the collaboration?

References

1. Blandford, A., Thimbleby, H., and Bryan-Kinns, N. Understanding interaction traps. In Proceedings of HCI 2003: Designing for Society – Volume 2. Research Press International, Bristol, UK, 2003, 57-60.
2. Cherubini, M., Venolia, G., DeLine, R., and Ko, A. J. "Let's Go to the Whiteboard: How and Why Software Developers Use Drawings". In CHI '07, pages 557–566.
3. Gurr, C. A. "Effective diagrammatic communication: Syntactic, semantic and pragmatic issues". Journal of Visual Languages and Computing, 10 (1999), 317-342.
4. Hutchins, E. (1987). Metaphors for interface design. ICS Report 8703. La Jolla: University of California, San Diego.
5. Larkin, J. H., and Simon. H. A. "Why a diagram is (sometimes) worth ten thousands words". Cognitive Science, 11, 1 (1987), 65-100.
6. McGookin, D., Brewster, S. "An Initial Investigation into Non-Visual Computer Supported Collaboration". Work in Progress In CHI'07.
7. Metatla, O., Bryan-Kinns, N., Stockman, T. "A Model for Structuring UML Class Diagrams to Support Non-Visual Interpretation and Navigation". In Proceedings of the 20th BCS HCI Conference , 2006.
8. Metatla, O., Bryan-Kinns, N., Stockman, T. "Using Hierarchies to Support Non-Visual Access to Relational Diagrams". In Proceedings of the 21st BCS HCI Conference , 2007.
9. Metatla, O., Bryan-Kinns, N., Stockman, T., "Constructing Relational Diagrams in Audio: The Multiple Perspective Hierarchical Approach". To appear in the proceedings of the 10th ACM Conference on Computers and Accessibility, 2008.
10. Metatla, O., Bryan-Kinns, N., Stockman, T., "Comparing Interaction Strategies for Constructing Relational Diagrams in an Audio-Only Interface". To appear in the proceedings of the 22nd British HCI Conference, 2008.
11. Scaife, M., and Rogers, Y. "External cognition, interactivity and graphical representations". In Proceedings of the IEE Colloquium: Thinking with diagrams. (London, UK, 1996). Digest No: 1996/010, 8/1-8/6.
12. Winberg. F., Bowers. J. "Assembling the Senses: Towards the Design of Cooperative Interfaces for Visually Impaired Users". Proceedings of the CSCW' 04, Chicago, USA.
13. Winberg, F. "Supporting Cross-Modal Collaboration: Adding a Social Dimension to Accessibility". Proceedings of the 1st Workshop on Haptic and Audio Interaction Design. HAID 2006. Glasgow, Scotland.
14. Zhang, J. "The Nature of External Representations in Problem Solving". Cognitive Science, 21, 2 (1997), 179-217.
15. Zhang, J. "A Representational Analysis of Relational Information Displays". International Journal of Human Computer Studies, 1996, 45, 59-74.

Social competence in schizophrenia

Mary Lavelle

Interaction Media and Communication group,
Computer Science Department,
Queen Mary University of London
maryl@dcs.qmul.ac.uk

1 Introduction

Schizophrenia is a severe mental illness affecting approximately 1% of the population. Although the causal factors remain somewhat obscure, it is the most disabling mental illness and people affected by it are among the most socially excluded in our society. Difficulties managing in social interaction are thought to be at the root of social exclusion. These interactional deficits are key because they influence every aspect of social integration, from maintaining friendships, to holding down a job, to living independently. Indeed, it may be that interactional problems are more important than symptoms or cognitive functioning in predicting long-term outcome in schizophrenia [1]. Currently we do not know what underlies these problems in interaction therefore they cannot be targeted therapeutically.

Early studies of human interaction proposed the theory of "interactional synchrony" [2][3]. Suggesting that synchrony occurs between the body movements of a speaker and a hearer within an interaction [2][3]. According to Kendon [3] interactional synchrony is a non-conscious phenomenon, which is seen predominantly at important junctures of an interaction, such as, at the beginning or end of an interchange. Interestingly, interactional synchrony was found to be either absent or greatly diminished in patients with a diagnosis of schizophrenia [2], suggesting difficulties with the subtle non-conscious aspects of interaction.

There is a growing body of work on social cognition attempting to explain the interactional difficulties displayed in schizophrenia. It has been shown that patients tend to have difficulties with social perception, in particular, nonverbal social cues [4][5][6]. However, these findings are based exclusively on experimental tests in laboratory settings. Typically, they require the patient to watch a video recording of other people interacting and then make judgements about the individuals in the video. There is a significant problem with the ecological validity of these studies. The dynamic, interactive and inter-personal demands of conversation are distinct from the skills required for relatively detached judgements of videos. Making judgements about actors is a different competence entirely from acting appropriately in an interaction and it is how patients interact in their daily lives is what matters to everyday functioning.

The aim of the present study is to examine the behaviour of patients with a diagnosis of schizophrenia in naturally occurring interactions. Specifically, we wish to investigate the non-conscious aspects of reciprocity within interaction such as; synchrony, mimicry and alignment. This study will focus on three main questions:

1. Can patients with a diagnosis of schizophrenia be discriminated from healthy controls?
2. What particular aspects of behaviour differ in this client group?
3. Are deficits exacerbated in multiparty interactions?

2 Method

Sample We will recruit thirty patients with a diagnosis of schizophrenia, fifteen age and sex matched healthy controls and 10 age and sex matched psychiatric controls, specifically patients who have been diagnosed as bipolar. Participants with anti-psychotic medication side effects will be excluded from the study.

Procedure participant are asked to complete two tasks. Firstly, a standardised interactional task that involves giving directions to another person [7]. People normally perform this task by moving their bodies to a position where they physically adopt the perspective of the person needing the directions (a particularly strong visible form of reciprocity) allowing comparison of the extent to which each sample group takes the confederates perspective. This task will be audio-visually recorded.

In the second task, participants are asked to produce a sketch map of the building for a visitor. This task, piloted by Healey [8], generates a good range of verbal and nonverbal interaction as the sketch map is negotiated and revised. This provides extensive opportunities for adopting and adapting other people's contributions. This task is initially dyadic as the participant is producing the map with one confederate. A short time after a second confederate joins and assists with the task, transforming the interaction from dyadic to multiparty. This task takes place in the Augmented Human Interaction lab and is motion captured using Vicon motion capture equipment. This allows us to capture a highly detailed 3D map of body movements during conversational interaction, enabling the data to be viewed from any given perspective.

Analysis Both verbal and nonverbal data from the interactions will be coded using ELAN annotation software.

References

1. Couture, M.S., Penn, D.L.: The functional significance of social cognition in schizophrenia: A review. Schizophrenia Bulletin. 32 (S1) pp.44-63 (2006)
2. Condon, W.S., Ogston, W.D.: Sound film analysis of normal and pathological behavior patterns. The Journal of nervous and mental disease. 143 (1) pp.338-347 (1966)
3.Kendon, A. Movement coordination in social interaction: Some examples described. Acta Psychologica. 32 pp 100-125 (1970)
4. Toomey, R., Wallace, C.J., Corrigan, P.W., Schuldberg, D., Green, M.F.: Social processing correlates of nonverbal social perception in schizophrenia. Psychiatry. 60 (4) pp. 327-340 (1997)

5. Toomey, R., Schuldberg, D., Corrigan, P., Green, M.F.: Nonverbal social perception and symptomatology in schizophrenia. Schizophrenia Research, 53 (1) pp. 83-91. (2002)
6. Wynn, J.K., Sergi, M.J., Dawson, M.E., Schell, A.M., Green, M.F.: Sensorimotor gating, orientating and social perception in schizophrenia. Schizophrenia Research. 73. pp.319-325. (2005)
7. Ono, T., Imai, M., Ishiguro, H.: A model of embodied communications with gestures between humans and robots. In: Proceedings of Twenty-third annual meeting of the cognitive science society, pp 732-737. Cogsci (2001)
8. Healey, P.G.T, Colman, M., Thirlwell, M. (2005) Analysing Multi-Modal Communication: Repair-Based Measures of Human Communicative co-ordiantion. In: Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems (2005)

Word Reordering in Statistical Machine Translation

Sirvan Yahyaei

Information Retrieval Group
Department Of Computer Science
Queen Mary, University of London
sirvan@dcs.qmul.ac.uk

1 Introduction

Assume we want to translate a foreign sentence $\mathbf{f} = f_1^J = f_1, ..., f_J$ into a target sentence $\mathbf{e} = e_1^I = e_1, ..., e_I$. The problem of statistical machine translation can be written as the following equation:

$$\hat{\mathbf{e}}(\mathbf{f}) = \arg\max_{\mathbf{e}}\{Pr(\mathbf{e}|\mathbf{f})\} \tag{1}$$

where $\arg\max$ is the search problem for finding the target sentence. Although, searching among all possible translations is an NP-Complete problem [6], the state-of-the-art SMT systems employ a set of features to model different aspects of the translation problem and use a dynamic programming approach to explore a part of search space and maximise the right hand side of equation 1. In this work, we are focusing on word reordering components of the SMT system.

Different languages have different word orders and as mentioned before trying all possible permutations is computationally intractable, so SMT decoders place restrictions to reduce the number of permutations. In the so-called IBM word-based models [1] only reordering of at most n words in a given time is allowed. Reordering based on the absolute and relative positions of the words are introduced in the IBM models. In phrase-based decoders, the next generation of the SMT decoders, reordering is allowed in a given window and according to the distance of jump the operation is penalized [7].

Recently, many SMT systems started to incorporate syntactic information to capture the word order differences between the languages. [2] limit the reordering to operations on syntactic parse-trees of source and target languages. Similarly, other approaches that use syntax trees of source or target languages rely on syntactic rules to apply the reordering operations. Although, syntax based decoders have recently shown a promising results [5], phrase-based decoders [7] are still more successful. In some approaches that have tried to employ syntactic information, transformation rules are applied to the source sentence to make it in an order similar to the target language. Transformation rules can be general syntax-based or specific lexicalised rules. Usually, in these approaches, source sentences of the training set are transformed and the reordered versions are used to learn the word alignments and phrases. [11] proposed a method to learn transformation rules from a parallel corpora. In their work, an algorithm is designed to extract re-write patterns, apply them to the source sentence and monotonically carry out the translation. At training time, to learn the rewrite patterns, source sentences are parsed, phrases are aligned and lexicalised and unlexicalised patterns are extracted. [4] present a similar approach to [11], but with hand crafted rules to re-write the source sentence. They argue that baseline phrase-based models are unable to perform the re-orderings such as those of between German and English. As they show, the main differences in German clause structure with English, it is clear that some of the re-orderings require long distance skips which is usually penalised very high by phrase-based decoder, that makes it almost impossible to occur.

In phrase-based decoders, typically the phrase table captures the local re-ordering between the words, however it fails to generalise it for unseen phrases and also long distance re-orderings. They mostly, rely on target language models to select among the different word orders. A language model is a statistical model that assigns a probability to a given sequence of words. In an n-gram language model, which are widely used in SMT decoders, the probability of generating a word depends on the previous $n-1$ words which are preceding it. The n-gram language model are only effective for short distance re-orderings and also [3] and [11] have shown that they are not enough to make all the reordering decisions.

2 Proposed Approaches

In lexicalised re-ordering models, a model is built to predict the word or phrase orientation during the decoding. These models assign a cost to the next candidate skip. The aim is to build a model that predicts the natural jump and penalise that jump less than other possible jumps by giving a lower cost to it. In this set of method, the models are mostly built based on word and phrase frequencies. [3] argue that n-gram language models are not enough to deal with even local re-orderings, thus they propose a distortion model to give a cost to each jump based on the words participated in the jump, The model computes the costs in word level, then combines the costs of the words to estimate the cost of the phrases. We propose two improvements over this work: Word clustering and phrase orientation from word distortion.

In word clustering, the idea is to classify the words based on their jump behaviours to address the issue of sparse data. Here we want to estimate the probability $Pr(d|f_i, f_j)$ where f_i and f_j are two foreign words and d is the jump we are about to make. We are looking for function $\mathcal{C}$ which maps every word f to their classes $\mathcal{C}(f)$. Now, we define the probability model$p$ as follows:

$$p(d|f_i, f_j, \mathcal{C}) := p(d|\mathcal{C}(f_i), \mathcal{C}(f_j)).p(f_i|\mathcal{C}(f_i)).p(f_j|\mathcal{C}(f_j)) \tag{2}$$

To find the optimum classes $\hat{\mathcal{C}}$, by performing maximum likelihood approach, we have:

$$\hat{\mathcal{C}} = \arg\max_{\mathcal{C}} p(d|f_i, f_j, \mathcal{C}) \tag{3}$$

d can be the exact distance (number of the words) between the two words, however to simplify the algorithm we can use one of the following alternatives:

- monotone and discontinuous
- swap, monotone and discontinuous
- discontinuous to left, swap, monotone and discontinuous to right

A variant of this approach is clustering the pair of words. Since words can have different grammatical roles in the sentence, by considering the translation of the word we reduce the ambiguity of it. In this case, equation 2 will be:

$$p(d|f_i, e_{a_i}, f_j, e_{a_j}, \mathcal{C}, A) := p(d|\mathcal{C}(f_i, e_{a_i}), \mathcal{C}(f_j, e_{a_j})).p(f_i|\mathcal{C}(f_i, e_{a_i})).p(f_j|\mathcal{C}(f_j, e_{a_j})) \tag{4}$$

There have been a lot of work on word clustering for different applications. In the context of statistical machine translation, usually words are clustered to overcome the issue of sparse data. [9] presented algorithms for classification of words in bilingual text. His implementation is widely used by SMT community. [8] gave algorithms for clustering bigrams and trigrams on mono-lingual text and [10] provided a different algorithm than [9] to cluster bilingual words.

Since we decode phrase by phrase, we need to convert the word distortion costs to phrase distortion costs. The aim is to make a bridge from word distortion model to phrase orientation model are: Firstly, lexicalised distortion model gives probabilities for jumps between words, but we need reordering models which work in phrase level. Secondly, to address the issue of unseen phrase pairs, we can use words inside those phrases to predict the orientation of the phrase. Following variations should be examined to find the best one:

- Last word of the previous phrase and the first word of the next phrase.
- Maximum / Minimum cost of the jump between the last word of previous phrase and all the words in the next phrase.
- Maximum / Minimum cost of the jump from all the words in previous phrase to the first word of next phrase.
- Maximum / Minimum cost between all pairs.
- Sum of all costs considering the alignments: Compute distortion costs for each phrase while building the model

References

[1] Peter F. Brown, Stephen Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. The mathematic of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311, 1994.

[2] Kenji Yamada and Kevin Knight. A decoder for syntax-based statistical mt. In *ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 303–310, Morristown, NJ, USA, 2001. Association for Computational Linguistics.

[3] Yaser Al-Onaizan and Kishore Papineni. Distortion models for statistical machine translation. In *ACL '06: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL*, pages 529–536, Morristown, NJ, USA, 2006. Association for Computational Linguistics.

[4] Michael Collins, Philipp Koehn, and Ivona Kučerová. Clause restructuring for statistical machine translation. In *ACL '05: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 531–540, Morristown, NJ, USA, 2005. Association for Computational Linguistics.

[5] Steve DeNeefe, Kevin Knight, Wei Wang, and Daniel Marcu. What can syntax-based MT learn from phrase-based MT? In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 755–763, 2007.

[6] Kevin Knight. Decoding complexity in word-replacement translation models. *Comput. Linguist.*, 25(4):607–615, 1999.

[7] Philipp Koehn, Franz Josef Och, and Daniel Marcu. Statistical phrase-based translation. In *NAACL '03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, pages 48–54, Morristown, NJ, USA, 2003. Association for Computational Linguistics.

[8] Sven Martin, Jörg Liermann, and Hermann Ney. Algorithms for bigram and trigram word clustering. *Speech Commun.*, 24(1):19–37, 1998.

[9] Franz Josef Och. An efficient method for determining bilingual word classes. In *Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics*, pages 71–76, Morristown, NJ, USA, 1999. Association for Computational Linguistics.

[10] Ye-Yi Wang, J. Lafferty, and A. Waibel. Word clustering with parallel spoken language corpora. *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, 4:2364–2367 vol.4, Oct 1996.

[11] Fei Xia and Michael McCord. Improving a statistical mt system with automatically learned rewrite patterns. In *COLING '04: Proceedings of the 20th international conference on Computational Linguistics*, page 508, Morristown, NJ, USA, 2004. Association for Computational Linguistics.

Health & Fitness project: A case study for application of Bayesian Network in modelling time risks in projects

Vahid Khodakarami

RADAR Group
Department Of Computer Science
Queen Mary, University Of London
vahid@dcs.qmul.ac.uk

Abstract. This presentation aims to demonstrate a Bayesian Network model for the infamous Health and Fitness Project (H&F) in the student Union building in Queen Mary University of London. Previously I developed a novel technique for capturing different aspects of risk and uncertainty in projects [Khodakarami et all 2007]. The model has several promising features that make it superior to available techniques in project risk management. The H&F project is used as a case study to illustrate how the model can improve the risk analysis process and capture a better picture of time related risks in the project.

1 Background

“Risk Management” has become an important part of “Project Management” and has attracted a wide range of research during the last decade [William 1995]. Since 1990 various risk management processes (RMP) have been proposed (PMBOK guide [PMI 2004], [PRAM 2004] and [RAMP 2005]). Apart from fundamental differences in assumptions and methodologies in these processes [Chapman 2006] they all aim to capture risk and uncertainty in the following three stages:

- Risk identification- attempts to distinguish the main sources of risk. This stage is also known as *qualitative risk management.*
- Risk Analysis- attempts to measure the risk and its impacts to different project outputs (i.e. cost, time, performance).
- Risk Respond- attempts to formulate management responses to the risk.

“Quantitative Risk analysis”, particularly the effect of risk and uncertainty on the project schedule is the centre of attention in this research. Having a Realistic Schedule for the project is one of the most cited factors of project success [Fortune & White 2006]. Despite availability of several tools and techniques, they often fail to capture uncertainty properly and produce inaccurate, inconsistent and unreliable results. [Khodakarami et all 2007] discuss why current techniques are not sufficient and what are missing?

- Concept of uncertainty in projects
- Subjectivity in project estimation
- Common causal factors (internally generated risks
- Trade-off between time, cost and performance
- Complex sensitivity analysis (what if?)
- Dynamic Learning

2 How Bayesian Networks can help

Bayesian network is a sufficiently well-defined language that can easily communicate to computers. They can also model deterministic, statistical and analogical knowledge in a meaningful model. This makes them suitable for a wide range of problems involving uncertainty and probabilistic reasoning. Bayesian Networks, as a powerful technique for decision support under uncertainty, have attracted a lot of attention in different fields. However their application in project risk management is novel. The key benefits of BNs that make them highly suitable for the project risk analysis domain are:

- Explicitly quantify uncertainty
- Make predictions with incomplete data
- Combine subjective and objective data
- Reason from effect to cause as well as from cause to effect
- Overturn previous beliefs in the light of new data(learning)

In my research I developed a novel model using Bayesian Networks that performs a CPM style scheduling [Khodakarami et all 2007]. The model has several promising features that enable us to model uncertainty in project time in a way that no other technique can. However finding empirical data for validating the model and comparing it with existing models was very challenging.

There are very few or no sets of case studies that would illustrate when the methods worked or failed. That is because: first, details of business projects and management methodologies and data are often considered proprietary. Second, national security projects may impose levels of classification on project details that effectively prohibit meaningful comparisons. Third, submitting to any evaluation of project risk management techniques and practices may reveal poor performance in either analysis or performance or both.

3 Case Study: Health and Fitness project at Queen Mary

Building a new Health and Fitness club for Queen Mary University of London was proposed and approved in 2005. The initiating and planning phases of the project started in 2006 and the actual work commenced in early 2007. The construction budget was set at 2.7m and the completion date set for August 2007. The project divided in two phases: I) Enabling work (demolishing the existing facilities including relocating the prayer room) and II) Main Scheme.

Phase I, started 22nd Jan 2007 and completed in line with the original plan. The relocation of prayer room suffered from 3-4 weeks delay. As a consequence, the main scheme and the completion date are affected.

The main scheme which started on 26th March, was originally estimated 23 weeks. In June 2007 it was reported that the project is 4 weeks behind the schedule. The progress was very slow and time risk was getting worse. It was obvious the completion date was not going to met. The delay expanded in July and August to 6 and 9 weeks respectively. In August (the original completion date!) the new completion date set at the end of October. Further risks and problems arose and consequently the project faced further and further delays. At the time of this report (25th Feb 2008) the project has not completed yet.

Although there was a complete project plan along with a comprehensive risk register, the project was well behind its planned time and budget. How useful were the project plan and the risk register? What went wrong and how it could be prevented? How Bayesian Networks can be used to model and analyse risks in this project. Details and results will be presented in the conference.

4 References

Chapman C. (2006). Key points of contention in framing assumptions for risk and uncertainty management. International Journal of Project management. Vol.24, p303-313.

Fortune J. and White D. (2006). Framing of project critical success factors by a system model. International Journal of Project management. Vol. 24, p53-65.

Khodakarami V, Fenton N and Neil M. (2007). Project Scheduling: Improved approach to incorporate uncertainty using Bayesian Networks. Project Management Journal. Vol. 38, No. 2, p39-49.

PMBOK (2004). A guide to Project Management Body of Knowledge. 3rd edition. Upper Darby, PA: Project Management Institute.

PRAM (2004). Project Risk analysis and Management Guide. 2nd edition. High Wycomb: Association for Project Management (APM) Publishing.

RAMP (2005). Risk Analysis and Management for Projects. Institute of Civil Engineering and the Faculty and Institute of Actuaries. London, Thomas Telford.

William T. (1995). A classified bibliography of recent research relating to project risk management. European Journal of Operational Research, Vol. 85, p18-38.