Portrait: Generating Personal Presentations

James Fogarty[†] Jodi Forlizzi[†*] Scott E. Hudson[†]

HCI Institute[†] and School of Design[*]
Carnegie Mellon University

Abstract

The rise of email and instant messaging as important tools in the professional workplace has created changes in how we communicate. One such change is that these media tend to reduce the presentation of an individual to a username, impacting the quality of communication. With current technology, including rich personal presentations in messages is still cumbersome. This problem is compounded by the fact that many of the potential benefits are realized by the recipient, though the sender incurs the costs.

This paper discusses the *Portrait* system, which demonstrates an automated approach to generating personal presentations for use in computer-mediated communication and other systems, such as awareness and ambient information displays. The Portrait system works by searching the web for photos or logos that represent individuals and organizations. It then combines these images to create personal presentations. By using the existing web presences of individuals and organizations, Portrait reduces the human costs of using pictures of people in communication and in information displays. In a small evaluation of this system, we found that it performed nearly as well as human searchers at the task of finding images for personal presentations.

Key Words: Ambient information displays, awareness systems, computer-mediated communications, email, World Wide Web.

1 Introduction and Motivation

As personal computers and the Internet have become key components of professional life, tools for computer-mediated communication have become important complements to the written document, the telephone conversation, and the face-to-face meeting. Email and instant messaging, in particular, have come to play an important role in professional interpersonal communication. While such media are often more convenient and cost-effective than other traditional communication media, they create interesting changes in how people communicate [23].

One such change is that these media tend to reduce the presentation of an individual to a username. The individual may select this username, or it may be determined by the technology. Regardless of how it is selected, the username carries less information about the individual than even the shortest face-to-face meeting. Although a written document is also not as rich as a face-to-face meeting, a letterhead does support the presentation of an identity. Organizations spend significant resources creating letterhead and other presentations of identify, probably because these artifacts have a significant impact on how the reader reacts to the content of the document. The formality present in business cards, letterheads, etc. is often absent from communication conducted via email and instant messaging.

In addition, a variety of ambient information displays use pictures of people to support the presentation of information [3, 9, 12]. One reason for this is that pictures allow easy recognition of who is related to a piece of information.

Because of their utility for communication systems and information displays, obtaining pictures of people is an important issue. One approach is to assume that each communication or piece of information includes the necessary picture. Another is to assume that the necessary picture is in a local database. Both of these approaches have their individual problems, but they share an important fundamental limitation. This fundamental limitation stems from the problem that many of the potential benefits of using pictures of people in communication and information displays are realized by the viewer of a display, but existing approaches generally impose the human costs on the person pictured. While many people make their picture available on their home page, they sometimes find it too cumbersome to make their picture available for other uses. This mismatch between the human costs of a system and the human benefits of a system can result in the failure of that system [10].

This paper presents *Portrait*, a system that automatically creates *personal presentations* for use with communications systems and information displays.

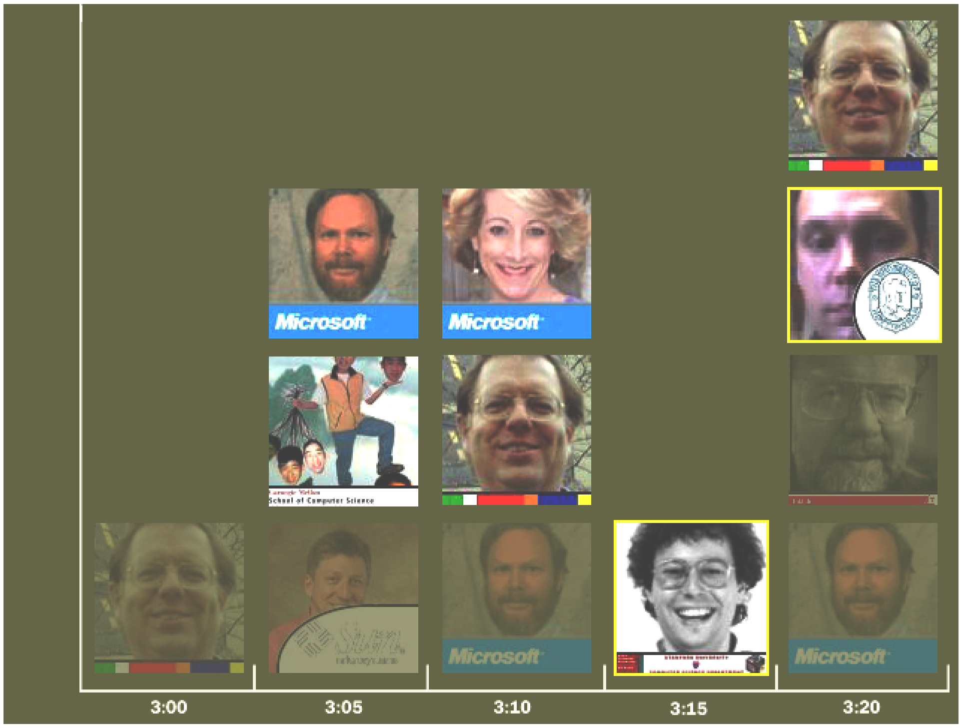

Figure 1. Personal Presentations used in a Design Sketch of an Ambient Email Display

Portrait searches the web for photos and logos to represent individuals and their organizational affiliations. Because many people and organizations have carefully constructed web presences, these images are often available on the home page for the person or organization. By reusing these images, we reduce the human costs of using pictures of people in communication systems and information displays.

We have been drawn to this problem by an interest in awareness systems that use glance techniques [4, 5, 11, 15, 22] and also by an interest in email glance displays [7, 13, 14]. These displays allow people to minimize the interruption of their current tasks while remaining aware of their communications and work context. The use of pictures of people in these displays can allow us to capitalize on our ability to quickly recognize faces. For example, an email glance display that uses pictures of people can allow us to quickly determine when we have received an email from a person important to us.

Apart from our interest in glance displays, there is evidence that trust development in computer-mediated communication can benefit from the inclusion of personal information [16]. Trust development in computer-mediated communication is an important issue for distributed organizations [17]. While large organizations can employ videoconferencing equipment to attempt to address this problem [2, 18], not all relationships justify the monetary costs and time overhead associated with this approach. This technology can also be problematic because it requires synchronous collaboration. It is, therefore, worth investigating approaches that are less expensive, both in terms of time and money [24].

Figure 1 presents some automatically generated personal presentations used in a design sketch of an ambient email display. These personal presentations were automatically generated during an evaluation discussed in this paper. This evaluation demonstrated Portrait's ability to find pictures of people with a success rate only slightly below the success rate of a human searcher.

Portrait currently searches for people using any combination of their name, email address, and organizational affiliation. The architecture, however, is pluggable and supports the creation of simple extensions that allow searches based on other information, such as an ID for an instant messaging system.

In the following sections, we present the pluggable architecture of Portrait, including details of our current search heuristics. We then present an evaluation of our current heuristics, followed by a discussion of some advanced concepts, related work, and a short conclusion.

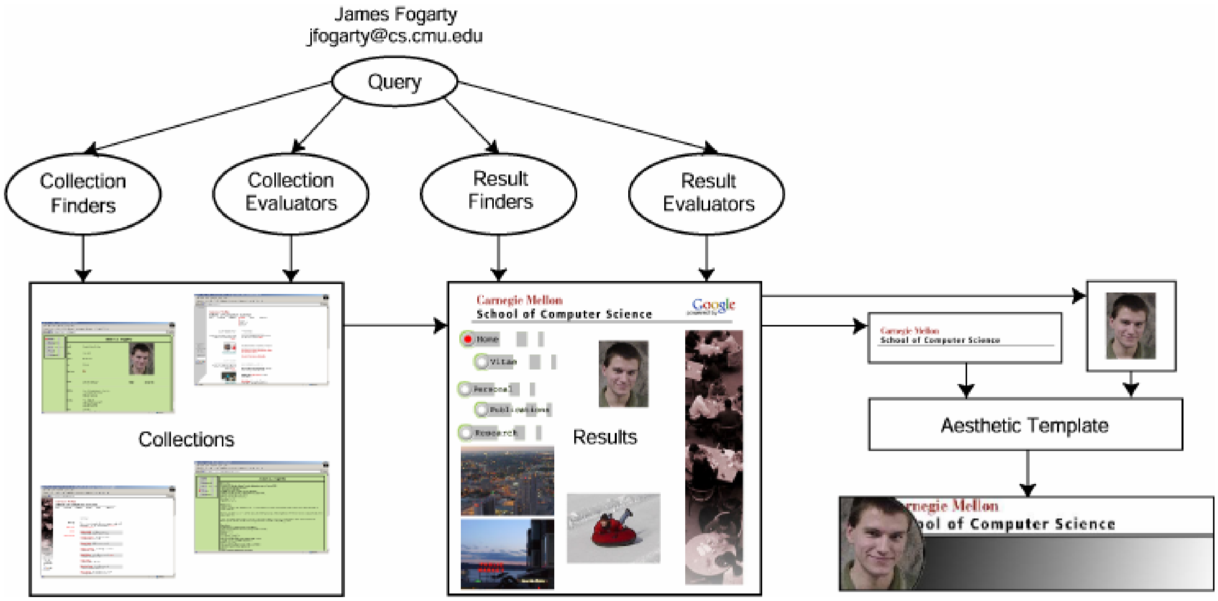

Figure 2. Portrait Architectural Diagram

2 Architectural Overview

Figure 2 presents the architecture of Portrait. The search for images to use in the creation of a personal presentation has two stages. First, Portrait searches for *collections* of images, which are typically web pages but might be any other source of images, including a database. Portrait then selects image *results* from relevant collections. This search process is based on four types of pluggable components. *Collection finders* are responsible for finding collections that might be relevant to a search. Next, c*ollection evaluators* examine the collections to estimate their relevance. *Result finders* then select results from the relevant collections. Finally, *result evaluators* examine the results to score their relevance.

When this search concludes, the best scoring images are selected and combined according to an *aesthetic template* [6]. These templates can provide sophisticated criteria for combining images and logos with fixed material in an aesthetically pleasing fashion. For example, they are capable of recognizing the positions of faces within images, then scaling and positioning those images within the template for the best effect.

Portrait uses weighted sums to combine the confidence levels indicated by each component related to a result. Each collection has a confidence level set by the collection finder that found it. For example, Portrait indicates more initial confidence in collections that were found by successfully guessing a standard URL format than in a collection that was found by following a series of links. Each collection evaluator also sets a confidence level for each collection. The confidence level of the collection is then used as an initial confidence level for any results selected from the collection. Results thus have initial confidence levels that are based on the context in which they are found. Result evaluators then provide evaluations that modify the initial confidence level of each result.

The components of Portrait interact using an event notification system. Events allow Portrait to keep track of the results with the highest confidence levels, allow other programs to monitor the progress of the search, and allow result finders to examine collections as they load. Event-based communication complements the multi-threaded implementation of Portrait. Loading each collection or result in its own thread prevents the system from being slowed by slow connections. To manage the number of active threads, we differentiate between threads that will load a collection, threads that will load a result, and threads related to event notification. We limit the active number in each thread category and manage new threads with a priority queue based on the confidence level of the component initiating a thread. This scheduling strategy ensures that threads with the most potential are not delayed by threads with less potential.

It is important to note that Portrait uses simple individual components. Our current heuristics, as discussed in the next section, are obviously naïve when considered individually. It is the combination of these individual heuristics that forms an effective system. Although we did not use a boosting approach to learn the relative weighting of our heuristics, our approach resembles the weighted combination of naïve components proven effective in boosting [7]. The success of our simple components can be partially explained by our two-stage architecture. Simple heuristics allow us to identify appropriate collections, and appropriate results can then be selected from these collections. For example, we can sometimes find the home page associated with an email address by simply adding www to the beginning of the email domain and putting a tilde before the email account. Similarly, following a URL in the email body or searching with Google [8] will often get us very close to an appropriate collection, thus drastically reducing the number of results that need to be considered by our result selection heuristics.

It is also important to note that simple individual components encourage the extension of Portrait. If Portrait were to be used with an instant messaging client, simple components could be added to search for collections related to the instant messaging ID of a person. Similarly, if Portrait were used in a banking environment, components could be added to encourage the selection of collections related to the banking industry. Finally, Portrait could benefit from simple components that prevent the selection of images from inappropriate web pages. These components might use a keyword method or might check a database maintained by a filtering service. A monolithic approach might not allow such extensions.

3 Current Heuristics

Because the Portrait system architecture is general and pluggable, it supports a wide range of possible search and evaluation heuristics. In our current implementation, we have focused on basic heuristics for creating presentations from the identification information available in email. Here we discuss the implemented heuristics.

3.1 Collection Finders

Collection finders find collections related to the current search. We currently use five types of collection finders:

1. *Guess Web Page URL*
   Guess Home Page From Email
   Guess Organization Page From Email
2. *Use URL found in Email Body*
3. *Google Search*
   Name, Email Address, Organization Name
4. *Follow Link by Content*
   Name, Email Address
   “Info”, “Personal”, “People”, “Pictures”
5. *Step Toward Root of Domain*

Some of these collection finders use knowledge very specific to the problem of finding pictures of people associated with an email address. Consider, for example, the collection finder that guesses URLs for the home page of a person from an email address. From the email address user@domain.com, the collection finder will guess the URL http://www.domain.com/~user and the URL http://www.domain.com/people/user, among others. If these guesses succeed in finding a page, there is a very high likelihood that the page is correct. This sort of very specific domain knowledge is not available to a general search engine, but yields good results when used in a specialized search like that conducted by Portrait. Further, the pluggable architecture of Portrait allows the easy inclusion of a variety of this sort of domain knowledge.

Note that our link-following collection finders follow links based on the textual content of the link, the URL of the link target, and the URL of an image used to present the link. Collection finders in the *Follow Link by Content* category select links containing synonyms related to a particular concept. For example, the collection finder for following “People” links follows links containing the text “people”, “faculty”, “member”, “staff”, “student”, or “employee”. We currently only look for keywords in English, but this same concept of synonyms would allow for following links in other languages.

3.2 Collection Evaluators

Collection evaluators determine confidence in collections. We currently use three types of collection evaluators:

1. *Web Page Title Contains Text*
   Name, Email Address, Organization Name
2. *Web Page URL Contains Text*
   Name, Email Address, Organization Name
3. *Web Page Content Contains Text*
   Name, Email Address, Organization Name

We currently differentiate between the appearance of text in the title of a page, the URL of a page, and the content of a page. For example, it is reasonable to expect that the appearance of an individual’s name in the title of a page is more meaningful than the appearance of the individual’s name in the content of the page. Similarly, a URL containing the user name of an email address is very likely to be related to that email address. We have already suggested that additional heuristics could be used to prevent selection of inappropriate web pages. Similar heuristics could be used to establish a preference for images found on web pages that appear more formal, discouraging such small mistakes as the selection of vacation pictures of the desired person.

3.3 Result Finders

Result finders use heuristic strategies to select appropriate images from within collections. We currently use five types of result finders:

1. *Contains Faces*
2. *Image Name Contains Text*
   "Logo", Organization Name
3. *Image Alt Text Contains Text*
   "Logo", Organization Name
4. *Image Links to Root of Domain*
5. *Generate Textual Presentation*

Our result finders for pictures of people use a robust system for detecting faces in images [19, 20]. This software package provides us with the location, the bounding box, and a confidence level for each appearance of a face in an image. We also use result finder that select images whose name or alt text contain a synonym of "Logo" or the organization name for which a search is being conducted. Because we have found that it is common to use an organizational logo to link to the top-level page in a domain, we have created a result finder that selects images used to link to the root of a domain. This is another example of very specific domain knowledge not available to a general search engine.

These result finders used are very restrictive compared to some possible result finders. For example, we might consider using a finder that selected logo images based on the presence of only a few colors in the image, which is a typical characteristic of logo images. However, button images on web pages also typically use only a few colors. We have opted for more restrictive result finders because we consider the selection of inappropriate images to be a more significant problem than failing to select an appropriate image.

In addition to our result finders that select result images, we have implemented result finders to create textual presentations. These allow us to support the case where we believe our search heuristics have failed, as indicated by the failure of the confidence level of the result to reach a specified threshold. When we find only a logo, for example, we combine it with a textual presentation of the individual. When we find only a picture, we combine it with a textual presentation of the organization. Our textual presentations are currently simple, but we intend to pursue the creation of more aesthetically interesting textual presentations as a part of future work. By creating textual presentations when the search process has failed, Portrait allows system designers to choose whether they want to handle every search the same or they want to use a different presentation format when the search process fails.

3.4 Result Evaluators

Result evaluators determine confidence in results. We currently use three types of result evaluators:

1. *Contains a Single Face*
2. *Image File Name Contains Text*
   Name, Email Address,
   "Logo", Organization Name
3. *Image URL Contains Text*
   Name, Email Address

Result evaluators are intended to differentiate between results selected from the same collection or from multiple collections with approximately equal confidence levels. As such, the confidence indicators they set have smaller magnitudes than the indicators set on collections. For example, image file names on web pages often contain the name of email user name of the person pictured in the email. This can be helpful, for example, when the image is one of several images found on a web page that provides pictures of all the members of a research group. Similarly, we use an evaluator that sets a high confidence indicator on images containing a single face, indicating a preference for pictures containing only the desired individual, as opposed to images containing multiple people.

4 Evaluation of Current Heuristics

To evaluate our current heuristics, we randomly selected 100 authors from the proceedings of a recent research conference. We queried Portrait for each author, using the author's name, email address, and organizational affiliation (or a subset of these parameters when they were not all available in the paper). We also paid two graduate students from another department to use the same information to manually search for the images. These students were instructed to use whatever methods they thought would help them find the images. Debriefings after the searches found that their preferred method was to use a search engine and inspect the pages returned by the search engine. (Figure 1 uses presentations generated during this evaluation).

Informal experimentation found a very low success rate for direct use of Google's image search. This is because the reasonable assumption that a human can select from several choices does not apply in the automated environment for which we have designed Portrait. This is also because the search technology used by Google cannot take advantage of the fact that we are looking for pictures of people. It commonly selected buttons or banners from personal home pages. Given its low success rate and the mismatch between the intended purpose of Portrait and Google's image search, we did not include it in our evaluation.

We considered Portrait's search for a picture to be successful when the highest scoring picture identified by the system was a picture of the correct person. We did not, therefore, consider as correct any images that contained more than one person. Some of the more interesting images that we did not consider correct included pictures of the research advisor of the person for whom we were searching, pictures of other people with the same name, and what appeared to be an avatar for use in 3D environments (although it was an avatar for the correct person).

We considered Portrait's search for a logo to be successful when the highest scoring logo identified by the system represented the top level of an organization (as in a university or a company) or an appropriate lower level of an organization (as in a research group within a company or a department within a university) with which the desired person was affiliated. We did not consider as correct any logos for lower levels of organizations with which the individual was not affiliated.

We considered nearly all images selected in our manual searches to be appropriate. This meant that some manually selected images were considered to be appropriate pictures even though they contained more than one person. This is reasonable in the context of a manual search because a picture containing only the desired person could be obtained by manually cropping the selected image. This also meant that some manually selected images were considered appropriate even though the image did not allow for the recognition of the person, as in the case where a person riding a bicycle was pictured from too far away to allow identification. We considered these images appropriate because they had been manually selected as representative images. There was exactly one case where we decided that a manually selected image was inappropriate; this was a case in which the manually selected picture was of a politician with the same name as the person who was the target of the search.

Table 1 summarizes the success rates of the searches. The success rates indicate what percentage of the 100 selected authors yielded an appropriate image in each search. Note that there are 25 cases where none of our searches found an appropriate picture, indicating it is likely that no picture is publicly available. Our heuristics for finding pictures of people performed slightly worse than the first manual case, but the difference between the two is not statistically significant ($p > .67$). The second manual case performed better than our heuristics ($p < .02$) and better than the first manual case ($p < .06$). We attribute the performance of the second manual case to exceptionally high diligence in spite of the tedious task. The searcher in the second case found pictures that were quite distant from the home page of the person, including cases where pictures were found on the web pages of social fraternities. We never expected our heuristics to find images this distant from a home page, and it is debatable whether it is appropriate to use images not intentionally made available on a home page. It is also worth noting that our heuristics succeeded in 5 instances where both manual searches failed. These cases include situations where the person had multiple home pages, not all of which contained a picture. The manual searches appear to have ended at a home page not containing a picture, while our heuristics selected a picture from one of the other home pages. This evaluation also indicated a need to improve our heuristics for finding logos. While color histogram and OCR techniques may be helpful in this regard, it will be important to avoid creating logo finders that mistakenly select button images and other inappropriate images.

	Picture	Logo
Any Search	75%	100%
Portrait Search	48%	79%
Manual Search 1	52%	98%
Manual Search 2	65%	99%

Table 1. Summary of Evaluation Success Rates

Finally, the sample used in this evaluation is biased toward the research community. We are comfortable with this bias because the systems that may benefit from Portrait currently exist primarily in the research community. As such systems become more common, it is also reasonable to expect that it will become more common for the people using these systems to have carefully constructed web presences. Further, additional heuristics could improve Portrait's performance on any particular population of people.

5 Using Specialized Knowledge

In situations where appropriate images are available at a defined location, Portrait should take advantage of this carefully captured knowledge. As an example, the internal network of the School of Computer Science at Carnegie Mellon University features an internal FaceBoard. For many members of the community, there is a picture available in the FaceBoard. Our architecture supports the use of this knowledge through a result finder component that encapsulates the functionality of the FaceBoard. This component responds only to queries related to an email address in the cs.cmu.edu domain. For queries from this domain, it uses knowledge of how to navigate the FaceBoard to check for a picture of the person. Because the organizational affiliation of the person is clear, this result finder also creates the appropriate logo result. This is another example of Portrait's extensible architecture supporting the use of domain knowledge that is not available to general search engines.

6 User Feedback to the Search Process

While programmers can easily extend Portrait, end-users are likely to become frustrated if it consistently selects the wrong image for a particular person. As a step towards addressing this problem, we have designed an approach to end-user criticism of the results and for end-user indication of appropriate results. When unhappy with the results, an end-user would access an application that lists people and the images used to present them. The end-user would indicate that a current picture or logo was inappropriate for a person, optionally providing the URL of a correct image.

This feedback process can be included without modifying existing components. When an end-user provides an image URL, we take an approach similar to the FaceBoard. A result finder component keeps a list of queries and provided URLs. When the result finder encounters these queries, it creates results from the provided URLs. When the end-user indicates that an image is inappropriate but does not provide the URL for a correct image, a result evaluator keeps a list of queries and results that are inappropriate for these queries. The result evaluator applies a large penalty to any results that the user has indicated are inappropriate for a query.

7 Related Work

Diogenes [1] is a system that determines whether or not pictures on the web are pictures of certain people. It could, for example, be used to find all web pages that contain a picture of a particular celebrity. It assumes a database is available that contains pictures of all the people against which the system should compare images encountered on the web. While both Diogenes and Portrait are related to searching for images of people, they address very different problems.

Ahoy! The Homepage Finder demonstrates the application of dynamic reference sifting to searching for homepages of individuals [21]. This system is particular interesting because it uses search histories to learn the location of web pages related to an organization. It can then try to automatically guess URLs within these locations when looking for a person associated with the organization.

8 Conclusion

We have presented Portrait, a system that automatically generates personal presentations from pictures of people and logos of their organizational affiliations. An evaluation indicates that our current heuristics perform the search only slightly worse than a graduate student. By reusing the existing web presences of people and organizations, our automated approach reduces the human costs of using pictures of people in communication and information displays.

Acknowledgements

We would like to extend special thanks to Henry Schneiderman and Takeo Kanade for making their face recognition software available for this project. We would like to thank Richard Ebert and Matthew Mowczko for manually searching for images during our evaluation. This work was funded in part by the National Science Foundation under Grant IIS-01215603 and the first author's NSF Graduate Research Fellowship. All trademarks are the property of their respective owners.

References

[1] Alp Aslandogan, Y., and Yu, C. T., "Diogenes: A Web Search Agent for Person Images", *Proceedings of the 2000 ACM International Conference on Multimedia.*

[2] Bos, N., Gergle, D., Olson, J. and Olson, G. Being there versus seeing there: Trust via video. *Extended Abstracts of the 2001 Conference on Human Factors in Computing Systems*.

[3] Cadiz, J., Venolia, G. D., Jancke, G., and Gupta, A. "Sideshow: Providing Peripheral Awareness of Important Information", *Microsoft Research Technical Report MSR-TR-2001-83*.

[4] Erickson, T., Smith, D., Kellog, W., Laff, M., Richards, J., and Bradner, E. "Socially translucent systems: social proxies, persistent conversation, and the design of Babble." *Proceedings of the 1999 Conference on Human Factors in Computing Systems.*

[5] Fish, R., Kraut, R., Root, R., and Rice, R. Video as a technology for informal communication. *Communications of the ACM*, 36, 1 (Jan 1993), Pages 48-61.

[6] Fogarty, J., Forlizzi, J. and Hudson, S.E., "Aesthetic Information Collages: Generating Decorative Displays that Contain Information", *Proceedings of the 2001 ACM Symposium on User Interface Software and Technology*.

[7] Freund, Y. and Schapire, R. "A Decision-Theoretic Generalization of On-line learning and an Application to Boosting", *Journal of Computer and System Sciences*, 55(1), pp. 119-139.

[8] *Google*, Web Page: http://www.google.com.

[9] Greenberg, S., and Rounding, M., "The Notification Collage: Posting Information To Public and Personal Displays", *Proceedings of the 2001 SIGCHI Conference on Human Factors in Computing Systems*.

[10] Grudin, J., "Why CSCW Applications Fail: Problems in the Design and Evaluation of Organizational Interfaces", *Proceedings of the 1988 Conference on Computer Supported Cooperative Work*.

[11] Gutwin, C., Roseman, M., and Greenberg, S. A usability study of awareness widgets in a shared workspace groupware system. *Proceedings of the 1996 Conference on Computer Supported Cooperative Work.*

[12] Hudson, S. E., and Smith, I., "Techniques for Addressing Fundamental Privacy and Disruption Tradeoffs in Awareness Support Systems", *Proceedings of the 1996 Conference on Computer Supported Cooperative Work*.

[13] Hudson, S. E. and Smith, I., "Electronic mail previews using non-speech audio. *Proceedings of the 1996 SIGCHI Conference on Human Factors in Computing Systems*.

[14] Ishizaki, S., "Multiagent model of dynamic design: visualization as an emergent behavior of active design agents", *Proceedings of the 1996 SIGCHI Conference on Human Factors in Computing Systems*.

[15] Lee, A., Girgensohn, A., and Schlueter K. NYNEX Portholes: Initial user reactions and redesign implications. *Proceedings of the 1997 International ACM SIGGROUP Conference on Supporting Group Work.*

[16] Moore, D., Kurtzber, T., Thompson, L., and Morris, M. Long and short routes to success in electronically mediated negotiations: Group affiliations and good vibrations. *Organizational Behavior and Human Decision Processes*, 77, 1, 1999.

[17] Olson, G., and Olson, J. Distance Matters. *Human Computer Interaction*, 15, 2000.

[18] Rocco, E. Trust breaks down in electronic contexts but can be repaired by some initial face-to-face contact. *Proceedings of the 1998 Conference on Human Factors in Computing System*.

[19] Schneiderman, H., "A Statistical Approach to 3D Object Detection Applied to Faces and Cars", *Doctoral Dissertation, Robotics Institute, Carnegie Mellon University* (Available as *Robotics Institute Technical Report 00-06*), May, 2000.

[20] Schneiderman, H., and Kanade, T., "A Statistical Model for 3D Object Detection Applied to Faces and Cars", *IEEE Conference on Computer Vision and Pattern Recognition*, June, 2000.

[21] Shakes, J., Langheinrich, M. and Etzioni, O. Dynamic reference sifting: A case study in the homepage domain. *Proceedings of the 1997 World Wide Web Conference*.

[22] Tang, J., Isaacs, E., and Rua, M. Supporting distributed groups with a montage of lightweight interactions. *Proceedings of the 1994 Conference on Computer Supported Cooperative Work.*

[23] Voida, A., Newstetter, W., and Mynatt, E.D. "When Conventions Collide: The Tensions of Instant Messaging Attributed", *Proceedings of the 2002 SIGCHI Conference on Human Factors in Computing Systems*.

[24] Zheng, J., Bos, N., Olson, J., and Olson, G. Trust without touch: Jump-start trust with social chat. *Extended Abstracts of the 2001 Conference on Human Factors in Computing Systems*.