An Interactive Computer Vision System
DyPERS: Dynamic Personal Enhanced Reality System

Bernt Schiele Nuria Oliver Tony Jebara Alex Pentland
bernt@media.mit.edu
Media Laboratory, Massachusetts Institute of Technology
Cambridge, MA 02139
http://www.media.mit.edu/vismod/demos/dypers

Abstract

DyPERS, 'Dynamic Personal Enhanced Reality System', uses augmented reality and computer vision to autonomously retrieve 'media memories' based on associations with real objects the user encounters. These are evoked as audio and video clips relevant for the user and overlayed on top of real objects the user encounters. The system utilizes an adaptive, audio-visual learning system on a tetherless wearable computer. The user's visual and auditory scene is stored in real-time by the system (upon request) and is then associated (by user input) with a snap shot of a visual object. The object acts as a key such that when the real-time vision system detects its presence in the scene again, DyPERS plays back the appropriate audio-visual sequence. The vision system is a probabilistic algorithm which is capable of discriminating between hundreds of everyday objects under varying viewing conditions (lighting, view changes, etc.). Once an audio-visual clip is stored, the vision system automatically recalls it and plays it back when it detects the object that the user wished to use to remind him of the sequence. DyPERS interface augments the user without encumbering him and effectively mimics a form of audio-visual memory. Performance is evaluated and usability results are shown.

1 Introduction

Research in computer vision has been focusing around the idea to create general purpose computer vision algorithms. The spirit of these algorithms has been manifested in Marr's book [Mar82] where vision is essentially defined as a reconstruction process which maps the visual data into representations of increasing abstraction. However, it has been realized that computer vision algorithms are only part of a larger system with certain goals and tasks at hand possibly changing over time. This observation has led to the emergence of research fields such as active [Baj85, Baj88], animate [Bal91], purposive vision [Alo90] as well as dynamic vision [Dic97]. Whereas the main concern of Marr's paradigm might be summarized as *generality* of computer vision algorithms, active vision research has been concentrated on the *adaptability* of algorithms directed by goals, resources and environmental conditions.

Using computer vision algorithms in the context of human computer interfaces adds at least one further criterium which we summarize as *usability*. Usability refers to the need to design algorithms in such a way that they can be used in a beneficial way in a human-computer interaction scenario. In other words, a computer vision algorithm is usable only if the human user gains an advantage in using the overall system. Even though this seems an obvious requirement it has deeper implications: first of all the system's response time has to be reasonable (ideally real-time). Furthermore, the system has to be robust and reliable enough in order to be usable in changing environments. On the other hand in a human-computer interaction scenario the user may assist the system to overcome limitations or to help bootstrap, if (and this is an *only if*) the user feels a benefit using the system.

In this paper we propose a system which uses computer vision in a human-computer interaction scenario. An advantage of human-computer interaction scenarios is that as we can actually enclose the human in the overall system loop. In order to do so the human has to be able to influence the system's behavior. In addition to this it is highly important, that the user obtains feedback from the system in order to understand the calculated results of the system. More specifically for system described here, the human uses a simple input device in order to teach the system. By observing the system's results he may understand limitations of the systems and may be able to assist the system in order to overcome them.

Obviously we do not want that the user adapts entirely to the system which is the case for traditional human-computer interfaces using only keyboard, mouse and screen. Furthermore, the user should not be obliged to know how the system works or even any implementation details. Rather, we are looking for scenarios where the user may benefit from using the system versus not using the system. Therefore, going along the discussion, we always have to keep in mind the usability of the system or, in other words, that future users of the system are only interested in a beneficial use of the system and not in the system in itself.

1.1 Motivation for DyPERS: A Dynamical and Personal Enhanced Reality System

As computation becomes widely accessible, transparent, wearable and personal, it becomes a useful tool to augment everyday activities. Certain human capabilities such as daily scheduling need not remain the responsibility of the user when they can be easily transfered to personal digital assistants. Certain tasks are excessively cumbersome to humans and involve little overhead computationally. An important one is memory. It is well-known that some things are best stored using external artifacts (such as handwritten or electronic notes) than in the human brain. However, it is also important to transfer these activities into a digital assistant in a natural, seamless way. Often, it is more cumbersome to encode the desired functionality into a computer than to manually perform it directly, in other words, to *transfer* something from reality into a virtual space. In such cases it is critical that the assistant operates autonomously without user intervention. DyPERS is a 'Dynamic Personal Enhanced Reality System' which addresses the above issues. It acts as an audio-visual memory assistant which reminds the user at appropriate times using perceptual cues as opposed to direct programming. The use of a head-mounted video camera and microphone mean that the user does not have to translate from their audiovisual environment into some computational language. DyPERS sees and hears what the user perceives, and continuously forms an audio-visual memory as a multimedia database that is available for subsequent playback. The user can then note which visual cues are important, causing DyPERS to learn how to recognize them (for future encounters), and associates them with the recorded memories.

When a cue is recognized at some later time, DyPERS automatically overlays these audio-video clips on the user's world through a heads-up-display (HUD)[FMS92], as a reminder of the content. This process is triggered when a relevant object is detected by the video camera system which processes visual data to recover meaningful correlations and associations with the memories using minimal user input.

2 Background and Related Work

This section describes related systems, compares them to DyPERS, and emphasizes the new contributions of DyPERS.

2.1 Memory augmentation

Some related memory augmentation systems are: Lamming's "Forget-me not" system ([LF93]), which is a personal information manager inspired by Weiser's ubiquitous computing paradigm, and the Remembrance Agent ([RSn96]), which is a text-based context-driven wearable augmented reality memory system.

2.2 Augmented Reality

Several augmented reality systems share some features with DyPERS. In [KVB97] a virtually documented environment system is described. The system is used to assist the user in some performance task. It registers synthetic multimedia data acquired using a head-mounted video camera in a similar way as DyPERS does, however information is retrieved explicitly by the user via a speech recognition system.

The direct precursor to DyPERS is described in [Lev97]. This system used machine vision to locate 'visual cues,' and then overlaid a stabilized image on top of the users view of the cue object. Visual cues were pre-stored images of objects that the user wanted to associate with some message or display. The machine vision algorithm used, however, was limited to 2D objects that were viewed from approximately vertically and head-on, and from approximately the same distance. Processing was done off-line. An earlier version of this system, described in [SMR+97], employed colored bar code tags as the visual cue.

Rekimoto et al. present in [RNg95] a system called NaviCam, which is a portable computer with a small video camera to detect pre-tagged objects. The system allows the user to view the real-world together with context sensitive information generated by the computer. NaviCam is extended in the Ubiquitous Talker ([RNg95]) and Ubiquitous Talker II (ShopNavi) ([NR96]), such that it incorporates a speech dialogue interface. The same authors have also developed a navigation systems called WalkNavi ([NR96]). Audio Aura ([MBWFd97]) is an active badge distributed system that augments the physical world with auditory cues. It allows passive interaction by the users who trigger the transmission of auditory cues as they move through their workplace. Finally Jebara et al. ([JEWv+97]) propose a vision-based wearable enhanced reality system – Stochasticks – for augmenting the billiards experience.

2.3 Perceptual Interfaces

Human-computer interaction has not fundamentally changed for nearly two decades. Most users are still limited to interacting with computers via keyboards and pointing devices. The bottleneck in improving the usefulness of interactive systems increasingly lies not in performing the processing task itself but in communicating requests and results between the system and its user ([JLMP93]). Faster, more natural and more convenient means for users and computers to exchange information are needed. On the user's side, interactive system technology is constrained by the nature of human communication organs and abilities; on the computer side, it is constrained by input/output devices and technologies. There has been much more emphasis in developing the computer-to-user direction of the communication –output devices– than its reciprocal direction from user-to-computer –input devices–. In consequence, any system or technology that improves the input channel would have a substantial positive effect in human-computer interaction.

In this sense and specially during the latest years there has been an increasing number of researchers in various areas of computer science developing technologies to add perceptual capabilities — such as speech, vision and touch — to human-computer interfaces. These *perceptual interfaces* are likely to be a major model for future human-computer interaction ([Tur97]).

From this perspective DyPERS is a *wearable perceptual interface*, an interface that closely matches the user's perceptual abilities. Wearable computing enables a unique first-person sensory viewpoint for the user interface. By incorporating sensors –camera and microphone– the amount of contextual information that is available to the wearable computer is vastly increased. One consequence is a more intelligent and fluid interface that uses the world as part of the interface, as opposed to just being limited to the desktop metaphor. Wearable computers offer a rich and new research field for computer-human interaction. Techniques from computer vision systems, indoor and outdoor position sensing [SKA97], voice systems [ES94], affect sensing [Pic97] and simple keyboard monitoring can be combined to minimize the need for direct user manipulation of the computer.

DyPERS incorporates new features that weren't present in any of the systems described above: multimedia user memory augmentation by using audio-visual associative memory, as opposed to the textual information (as in the Remembrance Agent). It also uses *generic real-time computer-vision object recognition* as opposed to more limited, pre-tagged solutions to identifying cue objects; on-line trainable audio-visual associative memory, as opposed to off-line pre-registered textual information and audio-visual output overlaid on reality as opposed to computer graphics or text-based messages.

Some other key properties of DyPERS that naturally arise from being an enhanced reality system are: augmentation or annotation of real objects with audio-visual information in a seamless way, personalization, adaptability, trainability, hands-free operation, mobility and perception (i.e. the computer's ability to perceive and act on attributes of its environment, enabling context dependent computing). This last feature has been called *environment-directed computing* [FKS97], where the wearable computer can automatically adapt to changes in the user's task, situation and context. DyPERS generic object recognition system is our approach to having a *situated computer* [HNBR97]. In other words, a computer system that is able to recognize specific objects in its environment – those that are relevant for the user – and augment them by displaying related audio-visual information. Situation awareness is of particular importance in wearable systems due to their intrinsic mobile nature. Wearable computers will potentially go everywhere, in a variety of situations where appropriate behavior for a given situation might be essential. We believe that the potential for wearable computers is best perceived not only as physical extensions of their users, but as mental and cognitive extensions. DyPERS, by associating audio-visual information to specific real objects, acts as a *memory augmentation* or *situated reminder* device.

3 System's overview

DyPERS building blocks are depicted in figure 1. The following describes the audio-visual association module, the generic object recognition algorithm used and gives a short overview of the hardware.

3.1 Audio-Visual Associative Memory system

The audio-visual associative memory module receives object labels along with confidence levels from the object recognition system. If the confidence is high enough, it will retrieve from memory the audio-visual information associated with the object the user is currently looking at and it will overlay this information on the real imagery that the user is perceiving.

The audio-visual recording module accumulates buffers containing audio-visual data. These circular buffers contain the past 2 seconds of compressed audio and video. Whenever the user decides to record the current interaction, the system stores the data until the user signals the recording to stop. The user moves his head mounted video camera and microphone to specifically target and *shoot* the footage required. Thus, an audio-video clip is formed. After recording such an audio-video clip, the user selects the object that should trigger the clip's playback which is done by directing the head-mounted video camera towards an object of interest and triggering the unit (i.e. pressing a button). The system then instructs the vision module to add the captured image to its database of objects and associate the

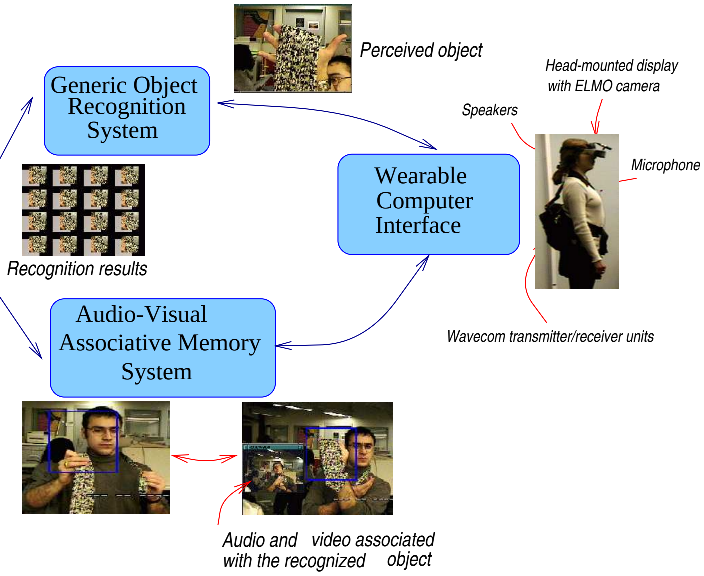

Figure 1: System's architecture

object's label to its most recent audio-visual clip. Additionally, the user can indicate negative interest in objects which might get misinterpreted by the vision system as trigger objects (i.e. due to their visual similarity to previously encountered trigger-objects). Thus, both positive and negative reinforcement can be performed in forming these associations. Therefore the user can actively assist the system to learn the differences between uninteresting objects and trigger objects.

The primary functionality of DyPERS can be projected on a simple 3 button interface (using a wireless 3-button mouse or a Toshiba Libretto notebook with a WaveLan connection): a record button, an associate button and a garbage button. The record button stores the A/V sequence. The associate button merely makes a connection between the currently viewed visual object and the previously recorded sequence. The garbage button associates the current visual object with a NULL sequence indicating that it should not trigger any play back. This helps resolve errors or ambiguities in the vision system which can quickly learn when it makes an error. This association process is shown in Figure 2. A very simple 3-command speech interface could also be incorporated following the same paradigm.

Whenever the user is not recording or associating, the system is continuously running in a background mode trying to find objects in the field of view which have been associated to an A/D sequence. DyPERS acts in consequence as a parallel perceptual remembrance agent that is constantly trying to recognize and explain – by remembering associations – what the user is paying attention to. Figure 3 depicts an example of the overlay process. Here, in the top figure, an 'expert' is demonstrating how to change the bag on a vacuum cleaner. The user records the process and then associates the explanation with the image of the vacuum's body. Thus, whenever the user looks at the vacuum (as in the bottom figure) he or she automatically sees an animation (overlaid on the left of his field of view) explaining how to change the dust bag.

3.2 Generic Object Recognition System

The input images sensed by the wearable camera are directly sent to the generic object recognition system. This system then tries to recognize the objects that the user is looking at. Upon recognition of some object it will send the recognition results – as object labels along with confidence levels – to the

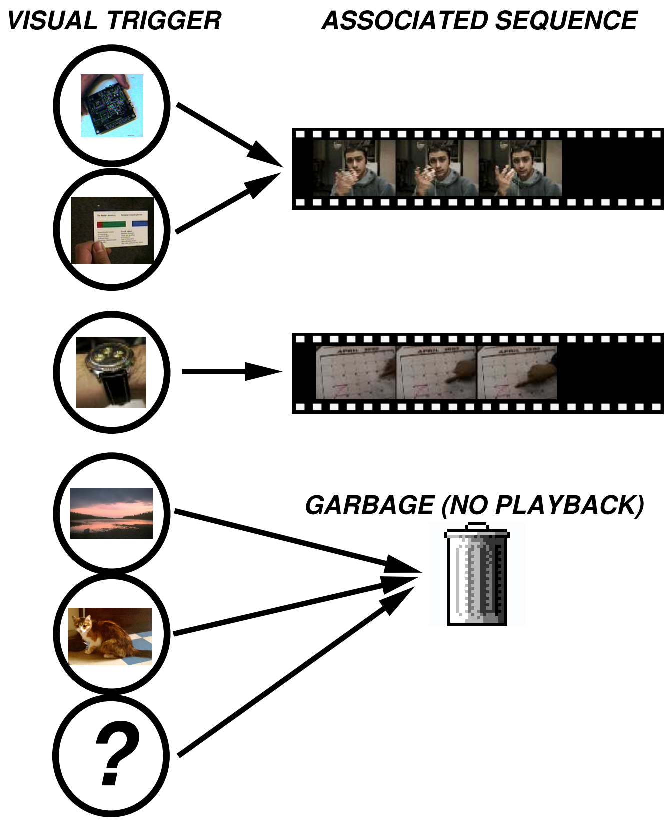

Figure 2: Associating A/V Sequences to Objects Recognized

audio-visual associative memory system.

The generic object recognition system used by DyPERS has been recently proposed by Schiele and Crowley [SC97]. A major result of their work is that a statistical representation based on local object descriptors provides a reliable means for the representation and recognition of object appearances. In the context of DyPERS this system is used to recognize previously recorded objects and to use the recognized objects as index into the audio-visual memory.

Schiele and Crowley [SC97, Sch97] presented a technique to determine the identity of an object in a scene using multidimensional histograms of vectors responses from local neighborhood operators. They showed that matching of such histograms can be used to determine the most probable object, independent of its position, scale and image-plane rotation. Furthermore they showed the robustness of the approach

Figure 3: Sample Output

to view points changes.

This technique has been extended to probabilistic object recognition [SC97], in order to determine the probability of each object in an image only based on multidimensional receptive field histograms. Experiments showed that only a relatively small portion of the image (between 15% and 30%) is needed in order to recognize 100 objects correctly. In the following we describe briefly the technique for probabilistic object recognition. The system runs at approximately 10Hz on a Silicon Graphics machine O2 using the OpenGL extension for real-time image convolution.

Multidimensional receptive field histograms are constructed using a vector of any linear filter. Due to the generality and robustness of Gaussian derivatives, we use multidimensional vectors of Gaussian derivatives (typically the magnitude of the first derivative and the Laplace operator at two or three different scales). In order to recognize an object we are interested in the calculation of the probability of the object O_n given a certain local measurement M_k (here a multidimensional vector of Gaussian derivatives). This probability $p(O_n|M_k)$ can be calculated by the Bayes rule:

$$p(O_n|M_k) = \frac{p(M_k|O_n)p(O_n)}{p(M_k)}$$

with $p(O_n)$ the a priori probability of the object O_n, $p(M_k)$ the a priori probability of the filter output combination M_k, and $p(M_k|O_n)$ is the probability density function of object O_n, which differs from the multidimensional histogram of an object O_n only by a normalization factor.

Having K independent local measurements $M_1, M_2, \ldots, M_K$ we can calculate the probability of each object O_n by:

$$p(O_n|M_1, \ldots, M_k) = \frac{\prod_k p(M_k|O_n)p(O_n)}{\prod_k p(M_k)} \quad (1)$$

M_k corresponds to a single multidimensional receptive field vector. Therefore K local measurements M_k correspond to K receptive field vectors which are typically from the same region of the image. To guarantee independence of the different local measurements we choose the minimal distance $d(M_k, M_l)$ between two measurements M_k and M_l sufficiently large (in the experiments described below we choose the minimal distance $d(M_k, M_l) \geq 2\sigma$).

In the following we assume the a priori probabilities $p(O_n)$ to be known and use $p(M_k) = \sum_i p(M_k|O_i)p(O_i)$ for the calculation of the a priori probability $p(M_k)$. Since the probabilities $p(M_k|O_n)$ are directly given by the multidimensional receptive field histograms, equation (1) shows a calculation of the probability for each object O_n based on the multidimensional receptive field histograms of the N objects. Perhaps the most tempting property of equation (1) is that we do not need correspondence. That means that the probability can be calculated for arbitrary points in the image. Furthermore the complexity is linear in the number of used image points.

3.3 Hardware

Currently, the system is fully tetherless with a wireless radio connection allowing the user to roam around a significant amount of space (i.e. a few office rooms). Plans for further evolving the system into a fully self-sufficient, compact and affordable form are underway. More powerful video processing in the PC104 platform and the VIA units would eventually facilitate it. However, for initial prototyping, a wireless system with wirelessly linked off board processing was acceptable. Figure 4 depicts the major components of DyPERS which are worn by the user during operation. The user dons a Sony GlassTron heads-up display with a semi-transparent visor. Attached to the visor is an ELMO video camera which is aligned

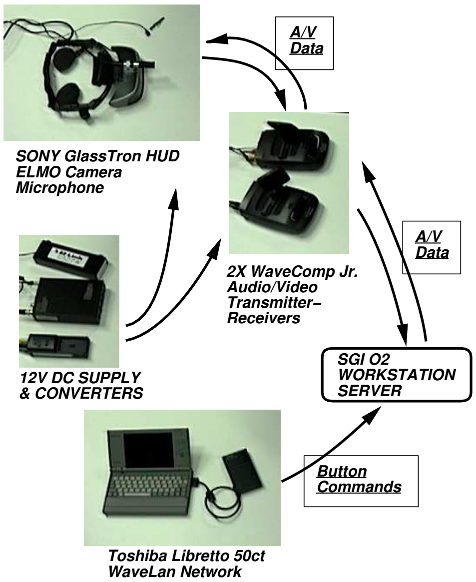

Figure 4: The Wearable Hardware System

as closely as possible with the user's line of sight[SMR+97]. Since the user has the option of viewing the world through the camera's point of view, a wide angle lens was preferred. In addition, a nearby microphone is also incorporated. The audio/video data captured by the system is continuously broadcast using a wireless radio transmitter. The main workstation receives this information, processes the audio and video streams and sends back the processed video and audio for output onto the GlassTron via a wireless radio receiver (on a channel different to the transmitter's channel). This wireless transmission connects the user and the wearable system to an SGI O2 workstation where the vision and other aspects of the system operate. The bidirectional audio-visual connection occurs in real-time (30Hz for the video). Note that it would be straightforward to port DyPERS code to a PC-based wearable platform with adequate computational power.

4 Scenarios

In this section we describe some interesting scenarios and applications that naturally fall into the record-and-associate paradigm of DyPERS. Four coarse categories are described with some mutual overlap. Some of the concepts have already been attempted successfully with the system and others remain to be investigated.

4.1 Recollection

One possibility is to use the system for remembering day-to-day information in an active setting. This could include daily scheduling and to-do list encoding. The user merely records his/her calendar or some notes indicating important things that need to be attended to. This recording can then be associated with the visual snapshot of the user's watch or a clock and would trigger playback whenever the user glanced at those objects. Alternatively, one could use the system to recollect a past interaction such as a communication with a business client and associate that clip with the client's business card. Whenever the user looked at the business card, the interaction would be replayed and important elements of the communication would be readily available.

4.2 Education

DyPERS has several interesting educational applications. These introduce an interesting variation to the system's usual operation since here the recordings are performed by an expert while the learner uses the system in playback mode. For instance, the expert could be an individual with knowledge of a foreign language (i.e. French), who would use DyPERS to record a variety of audio pronunciations of everyday objects and to associate them with visual snapshots of the objects. Thus, a novice French learner could hear the audio playback whenever facing an object of interest and hear the corresponding French phrase. Another scenario involves having a parent posing as an expert story teller or an entertaining baby sitter and a child as the novice. The adult could read a picture book and associate each picture with the audio on the same page. The child could then enjoy hearing a story which will be synchronized to the pages of a regular every day picture book.

4.3 Online Instruction: procedural information

Consider the completion of an activity or operation which involves many sequential steps and their corresponding actions. DyPERS could be trained by an expert to show a novice how to perform the complex activity online and interactively. At each landmark in the activity, the expert would record the next required sub-action (which would bring the user to the following state or landmark). For instance, consider the assembly of some pre-packaged furniture. The expert associates with the fully packaged item animated instructions on how to open the box and lay out the components. Subsequently, when the vision system detects the components placed out as instructed, it would trigger the corresponding assembly step. At each step, the system gives synchronized instructions about what to do next since the vision system is constantly tracking the evolution of the activity. In addition, if the novice performs an error and diverges from the instructions, the expert can train the system to detect this unusual state and show the user how to reverse out of this error and resume proper operation.

4.4 Augmented Perception

This category includes the variety of further sensory dimensions we may wish to incorporate to the inanimate objects we encounter. For instance, a compact disc could be associated with a small clip of the music it contains; a person with poor vision could benefit by listening to an audio description of the objects in his/her field of view; in virtual advertising one could associate everyday objects with sales pitch and in entertainment objects could be made come to life (i.e. a plant could ask to be watered). Ultimately, the visual appearance of an object can be augmented with further audio and video of relevant messages whereas the choice and content are left to the user's imagination.

5 Performance and Evaluations

Evidently, DyPERS has many different types of applications and it is unlikely to evaluate its performance in all possible situations. As has been pointed out in the introduction, the *usability* of the system is the most important criterium for the performance evaluation. Therefore, usability studies in certain controlled environments are insightful. Since the system features primarily audio-visual memory and significant automatic computer vision processing, our test conditions should evaluate these aspects in particular.

User Class	Description	Accuracy
A	With DyPERS	92.5 %
B	With Note Pad	83.75%
C	No Extra Paraphernalia	79.0%

Table 1: Subject Classes Accuracy

DyPERS was evaluated in a museum-gallery scenario. Audio augmented reality has been proposed by Bederson ([BH95]) in a similar museum situation. DyPERS expands the audio interface by providing personalized audio-visual augmented reality which is particularly suited for the museum situation. A museum constitutes a rich visual environment and is often accompanied with facts and details (from a guide or text) associated with the paintings. In addition, it is an educational experience that allows us to verify the system's usefulness as an educational tool.

A small gallery was created in our lab using 20 poster-sized images of various famous works ranging from the early 16th century to contemporary art. Three classes of users in different interaction modes (type A, type B and type C) were used in a walk through the gallery while a guide was reading a script that described the paintings individually. The guide presented biographical, stylistic and other information for each of the paintings while the subjects either used DyPERS (group A), took notes (group B) or simply listened (group C) to the explanations. Figure 5 shows a subject wearing DyPERS while listening to the guide during the gallery tour. After the completion of the guide's presentation, the subjects were required

Figure 5: A DyPERS user listening to a guide during the gallery tour

to take a 20-question multiple-choice test containing one query per painting presented. The questions ranged from date information, to stylistic questions to names of artists and works. In addition, the users had access to most of the images of the paintings since these were printed on the test sheets or still visible in the gallery. Thus, the subjects could refer back to these images while trying to answer the questions. For each test session, subjects of all three types described above were present and examined (i.e. A, B, and C were simultaneously present and, thus, variations in the guide's presentation do not affect their relative performance). Table 1 contains the accuracy results for each of the user groups. The results suggest that the subjects using DyPERS had an advantage over subjects without any paraphernalia or with traditional pencil and notepad. Moreover, the users wearing DyPERS reported the advantage of having the recorded data constantly available and triggered upon perception of each piece of art.

Currently, arrangements are being made with the List Visual Arts Center[1] for attempting the above testing in their publically accessible contemporary art gallery.

[1] List Visual Arts Center, 20 Ames Street, Media Arts and Sciences Laboratory, MIT, Cambridge, MA, 02139, USA.

6 Summary and Conclusions

In this paper, we have described DyPERS, a 'Dynamic Personal Enhanced Reality System' which uses computer vision and augmented reality to autonomously provide media memories relevant to real-world objects on a wearable computer. The system represents therefore an interactive computer vision system enclosing the human closely into the system's loop. We have described the three main building blocks of DyPERS, namely the wearable hardware and interface, the generic object recognition system and the audio-visual associative memory. We have also provided several application examples. Our preliminary experiments in a visual arts gallery environment suggest that the group using DyPERS would benefit of higher accuracy and more complete responses than any of the other groups (using paper notebook or no additional tool). These preliminary results are certainly encouraging even though considerable more work needs still to be done to establish the final usability and performance of DyPERS as a dynamic enhanced audio-visual memory system. We believe that augmented reality can enrich the interactions between people and their environment (including other people). Both auditory and visual augmented reality and perception play a fundamental role for building natural, seamless and intelligent interfaces.

Acknowledgments

Special thanks to Nitin Sawhney and Brian Clarkson for help with networking and audio. We also express our thanks to Pattie Maes and Thad Starner for their insightful comments. Thanks as well to all the subjects who participated in the experiment.

References

[Alo90] Y. Aloimonos. Purposive and qualitative active vision. In *Image Understanding Workshop*, pages 816–828, 1990.

[Baj85] R. Bajcsy. Active perception vs. pasive perception. In *IEEE Workshop on Computer Vision*, pages 55–59, 1985.

[Baj88] R. Bajcsy. Active perception. *Proceedings of the IEEE*, 296:996–1005, 1988.

[Bal91] D. Ballard. Animate vision. *Aritifcal Intelligence*, 48:57–86, 1991.

[BH95] B.B. Bederson and J.D. Holland. Human factors in computing systems and mdash. 1995.

[Dic97] E.D. Dickmanns. Vehicles capaple of dynamic vision. In *15th International Joint Conference in Artificial Intelligence*, 1997.

[ES94] L. Eric and C. Schmandt. Chatter: A conversational learning speech interface. In *AAAI Spring Symp. on Intelligent Multi-Media Multi-Modal Systems*, 1994.

[FKS97] S. Fickas, G. Kortuem, and Z. Segall. Software organization for dynamic and adaptable wearable systems. In *Intl. Symp. on Wearable Computers*, 1997.

[FMS92] S. Feiner, B. MacIntyre, and D. Seligmann. Annotating the real world with knowledge-based graphics on see-through head-mounted display. In *Proc. of Graphics Interface*, 1992.

[HNBR97] R. Hull, P. Neaves, and J. Bedford-Roberts. Towards situaded computing. In *Intl. Symp. on Wearable Computers*, 1997.

[JEWv+97] T. Jebara, C. Eyster, J. Weaver, T. Starner, and A. Pentland. Stochasticks: Augmenting the billiards experience with probabilistic vision and wearable computers. In *Intl. Symp. on Wearable Computers*, 1997.

[JLMP93] R.J.K. Jacob, J.J. Leggett, B.A. Myers, and R. Pausch. Interaction styles and input/output devices. *Behaviour and Information Technology*, 1993.

[KVB97] S. Kakez, C. Vania, and P. Bisson. Virtually documented environment. In *Intl. Symp. on Wearable Computers*, 1997.

[Lev97] J. Levine. Real-time target and pose recognition for 3-d graphical overlay. Master's thesis, EECS Dept., MIT, 1997.

[LF93] M. Lamming and Flynn. Forget-me-not:intimate computing in support of human memory. In *FRIEND21 Intl. Symp. on Next Generation Human Interface*, 1993.

[Mar82] D. Marr. *Vision*. W.H. Freeman and Company, 1982.

[MBWFd97] E.D. Mynatt, M. Back, R. Want, and R. Frederik. Audio aura: Light weight audio augmented reality. In *UIST*, 1997.

[NR96] K. Nagao and J. Rekimoto. Agent augmented reality: a software agent meets the real world. In *Proc. of Intl. Conf. on Multiagent Sys.*, 1996.

[Pic97] R. Picard. *Affective Computing*. MIT Press, 1997.

[RNg95] J. Rekimoto and K. Nagao. The world through the computer: computer augmented interaction with real world environments. *UIST*, 1995.

[RSn96] B. Rhodes and T. Starner. Remembrance agent: a continuously running automated information retrieval system. In *Intl. Conf. on the Practical Application of Intelligent Agents and Multi Agent Technology*, 1996.

[SC97] B. Schiele and J.L. Crowley. Recognition without correspondence using multidimensional receptive field histograms. Technical Report 453, MIT, Media Lab, 1997.

[Sch97] B. Schiele. *Object Recognition using Multidimensional Receptive Field Histograms*. PhD thesis, I.N.P.Grenoble, July 1997. English translation.

[SKA97] T. Starner, D. Kirsch, and S. Assefa. The locust swarm: an environmentally powered, networkless location and messaging system. In *Intl. Symp. on Wearable Computers*, 1997.

[SMR+97] T. Starner, S. Mann, B. Rhodes, J. Levine, J. Healey, D. Kirsch, R.W. Picard, and A.P. Pentland. Augmented reality through wearable computing. *Presence, Special Issue on Augmented Reality*, 1997.

[Tur97] M. Turk, editor. *Perceptual User Interfaces Workshop Proceedings*, 1997.