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

Bernt Schiele, Nuria Oliver, Tony Jebara, and Alex Pentland

Vision and Modeling Group
MIT Media Laboratory, Cambridge, MA 02139, USA
{bernt,nuria,jebara,sandy}@media.mit.edu
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 (view changes, lighting, 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. The DyPERS interface augments the user without encumbering him and effectively mimics a form of audio-visual memory. First results on performance and usability 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 like 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 the user feels a real 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 the 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 the user to adapt 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, 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.

In the following we will motivate the described system from an augmented reality point of view. Section 2 discusses related systems. Section 3 outlines the overall system while section 4 discusses the generic object recognition system. Section 5 lists several scenarios where a system like DyPERS might be used. Section 6 gives some experimental results of a museum tour scenario where DyPERS has been used to record and recall explanations of a guide about each painting.

1.1 Motivation for DyPERS: Dynamic 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 a user when they can be easily transfered to personal digital assistants. This is especially important for tasks that are excessively cumbersome to humans yet involve little computational overhead. An important one is memory or information storage. It is well-known that some things are better stored using external artifacts (such as handwritten or electronic notes) than in a human brain. However, it is also critical that the transfer of information to be processed (i.e. by a digital assistant) proceeds in a natural, seamless way. Often, it is more cumbersome for a user to input data and functionality into a computer than to manually perform a task directly. In other words, the *transfer* from reality into a virtual space is often too distracting to the user and reduces a digital assistant's effectiveness. In such cases it would be helpful that the assistant operates autonomously without user intervention. DyPERS is a 'Dynamic Personal Enhanced Reality System' which is motivated by 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. Using a head-mounted camera and a microphone, DyPERS sees and hears what the user perceives to collect a fully audio-visual memory. The resulting multimedia database can be indexed and played back in real-time. The user then indicates to DyPERS which visual objects are important memory cues such that it learns to recognize them in future encounters and associate them with the recorded memories.

When a cue is recognized at some later time, DyPERS automatically overlays the appropriate audio-video clip 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 constantly scans the visual field to detect objects which are associated with the memories.

2 Background and Related Work

This section describes related areas, compares other systems to DyPERS, and describes some new contributions emphasized by the proposed system.

Ubiquitous vs. Wearable Computing: Both wearable/personal computing and ubiquitous computing present interesting routes to augmenting human capabilities with computers. However, wearable computers attempt to augment the user directly and provide a mobile platform while ubiquitous computing augments the surrounding physical environment with a network of machines and sensors. Weiser [Wei91] discusses the merits of ubiquitous computing while Mann [Man97] argues in favor of mobile, personal audio-visual augmentation in his wearable platform.

Memory Augmentation: Memory augmentation has evolved from simple pencil and paper paradigms to sophisticated personal digital assistants (PDAs) and beyond. Some closely related memory augmentation systems include the "Forget-me not" system [LF93], which is a personal information manager inspired by Weiser's ubiquitous computing paradigm, and the Remembrance Agent [RS96], which is a text-based context-driven wearable augmented reality memory system. Both systems collect and organize data that is relevant to the human user for subsequent retrieval.

Augmented Reality: Augmented reality systems form a more natural interface between user and machine which is a critical feature for a system like DyPERS. In [KVB97] a virtually documented environment system is described which assists the user in some performance task. It registers synthetic multimedia data acquired using a head-mounted video camera. However, information is retrieved explicitly by the user via speech commands.

On the other hand, the retrieval process is automated in [Lev97], a predecessor of DyPERS. This system used machine vision to locate 'visual cues,' and then overlaid a stabilized image, messages or clips on top of the users view of the cue object (via a HUD). The visual cues and the images/messages had to be prepared offline and the collection process was not automated. In addition, the machine vision algorithm used, was limited to 2D objects viewed from head-on and at appropriate distance. An earlier version, described in [SMR+97], further simplified the machine vision by using colored bar code tags as the visual cue.

In [RN95] the NaviCam system is described as a portable computer with video camera which detects pre-tagged objects. Users view the real-world together with context sensitive information generated by the computer. NaviCam is extended in the Ubiquitous Talker [RN95] to include a speech dialogue interface. Other applications include a navigation system, WalkNavi [NR96]. Audio Aura [MBWF97] is an active badge distributed system that augments the physical world with auditory cues. Users passively trigger the transmission of auditory cues as they move through their workplace. Finally, Jebara [JEW+97] proposes a vision-based wearable enhanced reality system called Stochasticks for augmenting a billiards game with computer generated shot planning.

Perceptual Interfaces: Most human-computer interaction is still limited to keyboards and pointing devices. The usability bottleneck that plagues interactive systems lies not in performing the processing task itself but rather in communicating requests and results between the system and the user [JLMP93]. Faster, more natural and convenient means for users to exchange information with computers are needed. This communication bottleneck has spurred increased research in providing perceptual capabilities (speech, vision, haptics) to the interface. These *perceptual interfaces* are likely to be a major model for future human-computer interaction [Tur97].

3 System Overview

The system's building blocks are depicted in Figure 1. The following describes the audio-visual association module and gives a short overview of the hardware. The generic object recognition algorithm is described in section 4.

3.1 Audio-Visual Associative Memory System

The audio-visual associative memory operates on a record-and-associate paradigm. Audio-visual clips are recorded by the push of a button and then associated

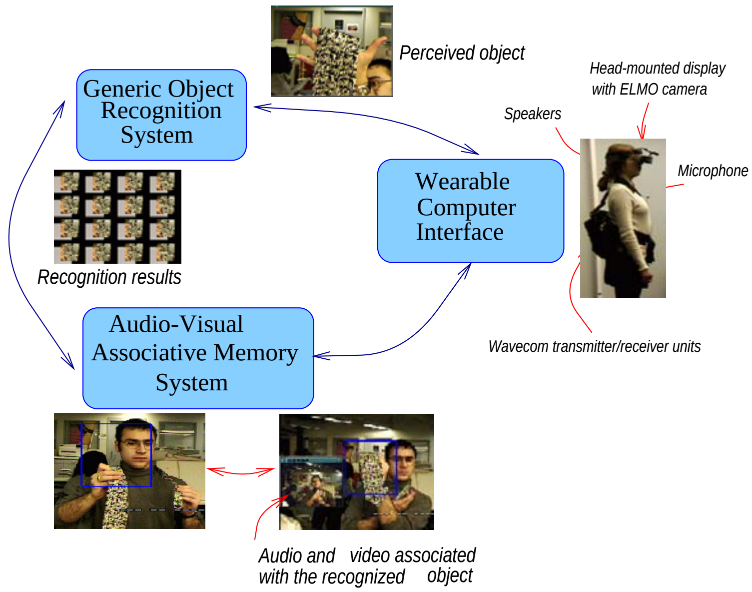

Fig. 1. System's architecture

to an object of interest. Subsequently, 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 retrieves from memory the audio-visual information associated with the object the user is currently looking at and overlays this information on the user's field of view.

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 a clip, the user selects the object that should trigger the clip's playback. This is done by by directing the 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 object's label to the most recently recorded A/V 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 used in forming these associations. Therefore the user can actively assist the system to learn the differences between uninteresting objects and important cue objects.

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/V sequence. DyPERS thus acts 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 2 depicts an example of the overlay process. Here, in the top of the 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 of the figure) he or she automatically sees an animation (overlaid on the left of his field of view) explaining how to change the dust bag. The recording, association and retrieval processes are all performed online in a seamless manner.

Fig. 2. Sample Output Through heads-up-display (HUD)

3.2 Wearable Computer Interface

As pointed out in the beginning, a system such as DyPERS has to be useful and usable by the person wearing it. Ideally we would like a non-intrusive system that did not require new infrastructure to be incorporated in the environment –such as tags, infrared transmitters, etc– and which can be used in a seamless way by its user.

Using a camera attached to the user's HUD and the generic real-time computer vision object recognition system described in section 4 DyPERS is able to perceive, identify and recognize the objects that the user is looking at. Using such a vision system circumvents many problems associated with tagging technologies, such as cost, size, range, power consumption and flexibility. From a perceptual viewpoint, DyPERS (in the same way as some other wearable systems [JEW+97,SMR+97,RN95]) sees what the user sees and hears what the user hears, being closer to the user's perception of the world.

The primary functionality of DyPERS is implemented in a simple 3 button interface (via a wireless mouse or a notebook PC with a wireless WaveLan). The user can select from a record button, an associate button and a garbage button. The record button stores the A/V sequence. The associate button merely

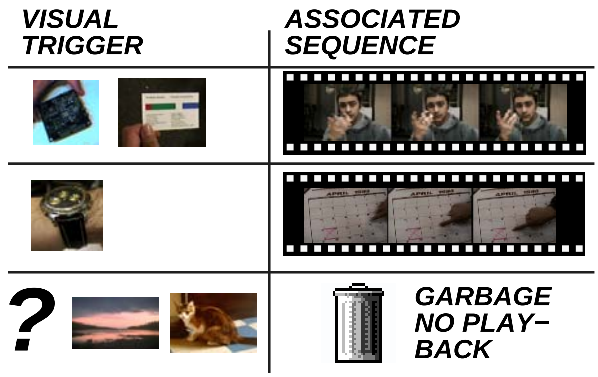

Fig. 3. Associating A/V Sequences to Objects

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. This association process is shown in Figure 3. In the current implementation of the system the interface is literately a three button interfaces. However, we are interfacing a small vocabulary speech recognizer in order to be replace the three buttons with spoken words.

3.3 Hardware

Currently, the system is fully tetherless with wireless radio connections allowing the user to roam around a significant amount of space (i.e. a few office rooms). Plans for evolving the system into a fully self-sufficient, compact and affordable form are underway. More powerful video processing in commercial units such as the PC104 platform or the VIA platform would eventually facilitate this process. However, for initial prototyping, a wireless system linked to 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 and headphones. Attached to the visor is an ELMO video camera (with wide angle lens) which is aligned as closely as possible with the user's line of sight [SMR+97]. Thus the vision system is directed by the user's head motions to interesting objects. In addition, a nearby microphone is incorporated. The A/V data captured by the camera and microphone is continuously broadcast using a wireless radio transmitter. 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 workstation collects the A/V data into clips, scans the visual scene using the object recognition system, and transmits the appropriate A/V clips back to the user. The clips are then rendered

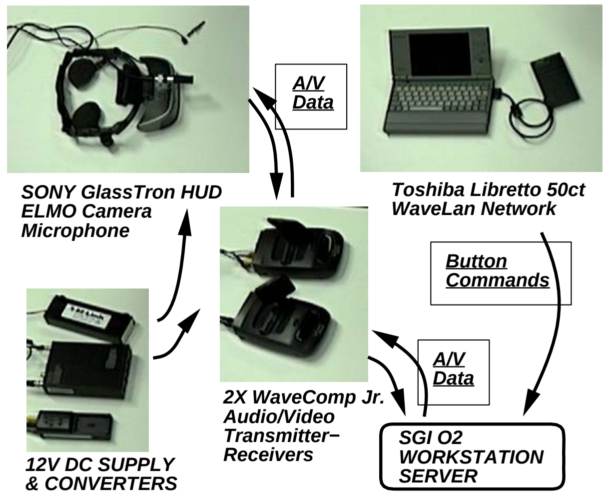

Fig. 4. The Wearable Hardware System

as an overlay via the user's GlassTron. Two A/V wireless channels are used at all times for a bidirectional real-time connection (user to SGI and SGI to user) [Man96].

4 Generic Object Recognition System

The video camera used by DyPERS is aligned with the line of sight of the user (see figure 8). Therefore, by gazing at interesting objects, the user directs the input to the recognition system which tries to recognize previously recorded objects. The recognition results are then sent to the audio-visual associative memory system which plays the appropriate clip.

The generic object recognition system used by DyPERS has been recently proposed by Schiele and Crowley [SC96]. 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.

Objects are represented by multidimensional histograms of vector responses from local neighborhood operators. Figure 5 shows two examples of two-dimensional histograms. Simple matching of such histograms (using χ^2-statistics or intersection [Sch97]) can be used to determine the most probable object, independent of its position, scale and image-plane rotation. Furthermore the approach is considerably robust to view point changes. This technique has been extended to probabilistic object recognition [SC96], in order to determine the probability of each object in an image only based on a small image region. Experiments (briefly described below) showed that only a small portion of the image (between 15% and 30%) is needed in order to recognize 100 objects correctly. In the following we summarize the probabilistic object recognition technique used. The current system runs at approximately 10Hz on a Silicon Graphics O2 machine using the OpenGL extension library for real-time image convolution.

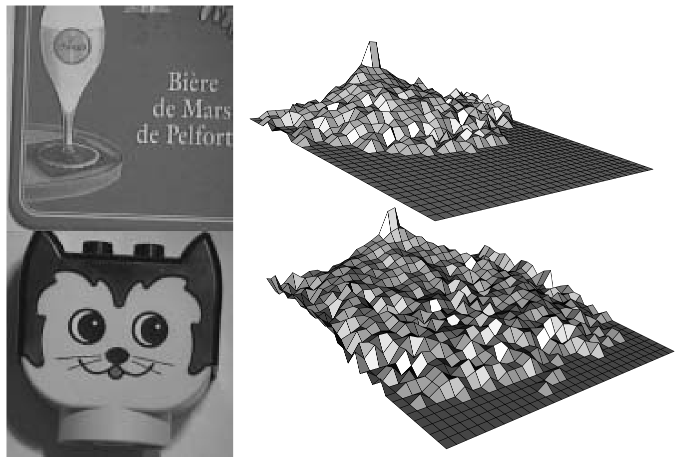

Fig. 5. Two-dimensional histograms of two objects corresponding to a particular viewpoint, image plane rotation and scale. The image measurement is given by the Magnitude of the first derivative and the Laplace operator. The resolution of each histogram axis is 32.

Multidimensional receptive field histograms are constructed using a vector of arbitrary linear filters. Due to the generality and robustness of Gaussian derivatives, we selected multidimensional vectors of Gaussian derivatives (e.g. the magnitude of the first derivative and the Laplace operator at two or three different scales).

It is worthwhile to point out that the object representation is very general and can be used for a wide variety of objects. The objects most suited for the representation contain enough local texture and structure to be coded by the multidimensional histograms. A useful feature of the recognition system is that it often matches visually similar objects such as two business cards from the same company. In order to discriminate these cards a more specific system such as a character recognition system should be used. Since the response time of the system is only in the order of 100ms we are planning to use the result of the system to trigger more specific recognition systems as appropriate.

4.1 Probabilistic Object Recognition

In order to recognize an object, we are interested in computing 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 using 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
- $p(M_k|O_n)$ 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:

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

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 to be sufficiently large (in the experiments below we chose the minimal distance $d(M_k, M_l) \geq 2\sigma$).

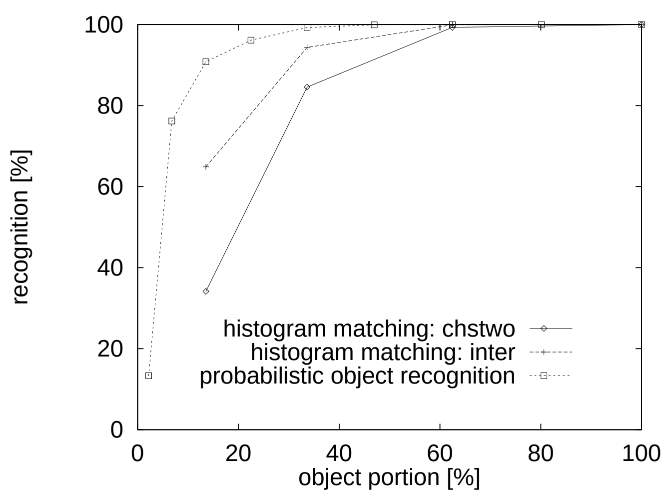

Fig. 6. Experimental results for 103 objects. Comparison of probabilistic object recognition and recognition by histogram matching: χ^2_{qv} (chstwo) and $\cap$ (inter). 1327 test images of 103 objects have been used.

Fig. 7. 25 of the 103 database objects use in the experiments.

In the following we assume all objects to be equally probable: $p(O_n) = \frac{1}{N}$ with N the number of objects. We 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 remarkable property of Equation (1) is that no correspondence needs to be calculated. That means that the probability can be calculated for arbitrary points in the image. Furthermore the complexity is linear in the number of image points used.

Equation (1) has been applied to a database of 103 objects. Figure 7 shows some of the objects used. In an experiment 1327 test images of the 103 objects have been used which include scale changes up to $\pm 40\%$, arbitrary image plane rotation and view point changes. Figure 6 shows results which were obtained for six-dimensional histograms, e.g. for the filter combination $Dx - Dy$ (first Gaussian derivatives in x- and y-direction) at three different scales ($\sigma = 2.0$, $= 4.0$ and $= 8.0$). A visible object portion of approximately 62% is sufficient for the recognition of all 1327 test images (the same result is provided by histogram matching). With 33.6% visibility the recognition rate is still above 99% (10 errors in total). Using 13.5% of the object the recognition rate is still above 90%. More remarkably, the recognition rate is 76% with only 6.8% visibility of the object. See [SC96,Sch97] for further details.

5 Scenarios

This section briefly describes some applications of DyPERS using the record-and-associate paradigm:

– Daily scheduling and to-do list can be stored and associated with the user's watch or other personal trigger object.
– An important conversation can be recorded and associated with the individual's business card.
– A teacher records names of objects in a foreign language and associates them with the visual appearance of the object. A student could then use the system to learn the foreign language.
– A story teller could read a picture book and associate each picture with its text passage. A child could then enjoy hearing the story by triggering the audio clips with different pages in the picture book.
– The system could be used for online instructions for an assembly task. An expert associates the image of the fully packaged item with 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 triggers the subsequent assembly step.
– A person with poor vision could benefit by listening to audio descriptions of objects in his field of view.
– The visual appearance of an object can be augmented with relevant audio and video or messages. For instance, the contents of a container could be virtually exposed after it is sealed.

Many of the listed scenarios are beyond the scope of this paper. However, the list should convey to the reader the practical usefulness of a system such as DyPERS. In the following we describe one application in further depth and show test results.

6 A Sample Test Scenario

Evidently, DyPERS has many applications and it is unlikely to evaluate its performance in all possible situations. A *usability* study in a sample environment was selected to gain insight on real-world performance of the system as a whole. Since the system features audio-visual memory and significant automatic computer vision processing, test conditions involved these aspects in particular.

DyPERS was evaluated in a museum-gallery scenario. Audio-only augmented reality in a museum situation was previously investigated by [Bed95]. The museum constitutes a rich visual environment (paintings, sculptures, etc.) which is accompanied by many relevant facts and details (usually from a guide or text). Thus, it is an audio-visual educational experience and well-suited for verifying 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 human participants (types A, B, and C) were tested in a walk-through of the gallery while a guide was reading a script describing the paintings. The guide presented biographical and stylistic information about each painting while the subjects either used DyPERS (group A), took notes (group B) or simply listened attentively (group C). The subjects knew that they would be tested after the tour. Figure 8 shows a subject wearing DyPERS while listening to the museum guide.

Fig. 8. A DyPERS user listening to a guide during the gallery tour

After the completion of the tour, the subjects were given a 20-question multiple-choice test containing one query per painting presented. In addition, the users had visual access to the paintings since these were printed on test sheets or still visible in the gallery. Thus, the subjects could refer back to the images while being tested. 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 standard pencil and paper notes. Currently, arrangements are being made with the List Visual Arts Center[1] for attempting the above test in their publically accessible contemporary art gallery.

[1] 20 Ames Street, MIT, Cambridge, MA 02139

Group	Description	Range	Average Accuracy
A	DyPERS	90%-95%	92.5 %
B	Note Pad	75%-95%	83.8%
C	No Aid	65%-95%	79.0%

Table 1. Subject Classes Accuracy

7 Summary and Conclusions

We have introduced an interactive computer vision system called DyPERS ('Dynamic Personal Enhanced Reality System'). The system combines computer vision and augmented reality to autonomously provide media memories related to real-world objects via a wearable platform. It allows a user to collect audio-visual clips in a seamless way and to retrieve them for playback automatically and meaningfully. The generic object recognition system has been described and its performance characteristics have been summarized. We have also discussed the two remaining building blocks of the system, namely the wearable hardware and interface, and the audio-visual associative memory. In addition, several application examples that DyPERS could span were enumerated.

Experiments in a visual arts gallery environment suggest that subjects using the computer vision system DyPERS would benefit of higher accuracy and more complete responses than participants using paper notes or no tools for information retention. These preliminary results are encouraging although more work is being planned to establish a final usability and performance evaluation. Nevertheless, the platform does provide interesting arguments for ways augmented reality and artificial perception can enrich the user and play a fundamental role in building a natural, seamless and intelligent interface.

Acknowledgments

Thanks to the participants involved in the experiment and to Nitin Sawhney, Brian Clarkson, Pattie Maes and Thad Starner for help and comments.

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.

[Bed95] B.B. Bederson. Audio augmented reality: A prototype automated tour guide. In *ACM SIGCHI*, pages 210–211, 1995.

[Dic97] E.D. Dickmanns. Vehicles capaple of dynamic vision. In *15th International Joint Conference in Artificial Intelligence*, 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*, pages 78–85, 1992.

[JEW+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*, 12(2):69–70, 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, June 1997.

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

[Man96] S. Mann. Wearable, tetherless computer-mediated reality. Technical Report 361, M.I.T. Media Lab, 1996.

[Man97] S. Mann. Wearable computing: A first step toward personal imaging. *IEEE Computer; http://wearcam.org/ieeecomputer.htm*, 30(2), February 1997.

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

[MBWF97] E.D. Mynatt, M. Back, R. Want, and R. Frederik. Audio aura: Light weight audio augmented reality. In *UIST*, pages 211–212, 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.

[RN95] J. Rekimoto and K. Nagao. The world through the computer: computer augmented interaction with real world environments. *UIST*, pages 29–36, 1995.

[RS96] 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*, pages 487–495, 1996.

[SC96] B. Schiele and J.L. Crowley. Probabilistic object recognition using multidimensional receptive field histograms. In *13th Intl. Conf. on Pattern Recognition, Volume B*, pages 50–54, August 1996.

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

[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.

[Wei91] M. Weiser. The computer of the twenty-first century. *Scientific American*, 1991.