# MEDIABEADS: AN ARCHITECTURE FOR PATH-ENHANCED MEDIA APPLICATIONS

*Michael Harville, Ramin Samadani, Dan Tretter, Debargha Mukherjee, Ullas Gargi, Nelson Chang*
Hewlett-Packard Laboratories, Palo Alto, CA 94304 USA

## ABSTRACT

Tagging digital media, such as photos and videos, with capture time and location information has previously been proposed to enhance its organization and presentation. We believe that the full path traveled during media capture, rather than just the media capture locations, provides a much richer context for understanding and "re-living" a trip experience, and offers many possibilities for novel applications. We introduce the concept of *path-enhanced media*, in which media is associated and stored together with a densely sampled path in time and space, and we present the *MediaBeads* architecture for capturing, representing, browsing, editing, presenting, and searching this data. The architecture includes, among other things, novel data representations, new algorithms for automatically building movie-like presentations of trips, and novel search applications.

## 1. INTRODUCTION

With the rapid growth of location-determining technologies such as those based on the Global Positioning System (GPS), it is likely that, in the near future, and without manual intervention, accurate location and time information can be measured by many consumer digital media recording devices, and applied as tags to captured media such as photos, videos, and audio recordings. A number of researchers have explored the use of spatiotemporal tags for enhancing the organization and consumption of digital media. For example, Toyama et. al. [14] present an extensive investigation into creating, organizing, and interacting with geo-referenced image databases. Keranen et. al. [5] describe a user interface that allows sharing and browsing of multimedia information with a location map context in an online mobile community. Several systems for attaching digital media to locations on maps have been described [1, 4, 8, 12], and the large community of Geographic Information Systems (GIS) researchers have sometimes considered spatiotemporally tagged media [9]. In all of this work, it is demonstrated that location and time gives *context* to media that is often not obvious from the media content itself, and that facilitates its relation to other media and other data organized by time and/or location.

Spatiotemporal media tags lack, however, in their ability to convey a "story" associated with a collection of media. In many application contexts, such as a family vacation, the captured media represents only part of the significant events, landmarks, and people encountered. Even if placed on a map or presented in time order of capture, the media alone provides a disjointed and incomplete view of the trip, since it is difficult and distracting for people to record everything of interest. One's judgement of what was "interesting" may also evolve long after the trip has concluded. Hence, when attempting to tell the story of a trip to someone who was not present, people must usually "fill in the blanks" between their photos, video, and audio to explain other events that occurred during the trip but are not represented in the media, and they must remember to emphasize transitions from one important location or event to the next.

We believe that *path* data - that is, the route in space and time traveled by the capturer of the media - provides a wealth of contextual information that links isolated media files, facilitates the creation and telling of stories from collections of media, and provides novel


**Fig. 1**. Map-based view of path-enhanced media for a trip through San Francisco. Path is in blue, with icons indicating captured media, and arrows indicating capture device location and orientation.

avenues for enhancing, analyzing, and leveraging a trip using spatiotemporally organized databases. Advantages of path context, as shown in Figure 1, over isolated spatiotemporal media tags include:

- **Improved recall:** When overlaid on a map, the traveled trip path may spark remembrances not represented in the captured media.
- **Realistic re-living of the journey:** Movement along a path can be "played" by an animation, triggering display of media when the animation reaches their capture locations. This provides a more immersive re-experiencing of the "journey" that the media is about.
- **More annotation opportunities:** Path data provides anchors to which media from other sources (such as stock photo databases), or other types of data (such as voice narrations, or restaurant or hotel names), can be attached, even when and where no media was captured. This allows for embellishment of multimedia trip presentations that is particularly important at locations for which people were unable to record media or later wish they had done so.
- **Powerful new queries:** Questions such as "Am I in someone else's picture?" are difficult to answer if the locations of people are known only at the instants when they capture media. In contrast, if two people record paths along with their media, photo capture times and locations of one person may be compared with all path points of the other, not just those where the latter took pictures. Further, one may answer questions like "Did someone take a photo of the event I was at but was unable to capture?", and "Who took a similar trip as me at about the same time?".

In this paper, we introduce the concept of *path-enhanced media*, or *PEM*, in which digital media is stored with and organized around a densely sampled path in time and space. We describe methods for

representing and storing such data, and then discuss our *MediaBeads* architecture for creating, browsing, editing, augmenting, presenting, and querying it. We describe key algorithmic components of the architecture, and enumerate many new capabilities it enables beyond what is possible with simple spatiotemporal media tags. A companion paper [10] details prototype systems for editing, presenting, and interacting with PEM data, and presents experimental results. These systems, along with an experiment in creating an interactive DVD presentation from PEM data, are summarized briefly here.

## 2. PATH-ENHANCED MEDIA

Path-enhanced media consists essentially of two components: path data, and digital media. We represent a path as a time-ordered series of *geo-temporal anchors*, or GTAs, each of which is a point in space and time with optional links to media files whose capture was in progress at that place and time. When information such elevation, device orientation (e.g. compass heading and/or tilt with respect to the ground normal), device imager focal length and/or field-of-view (FOV) angle, and temperature are also measured, we augment the GTAs with this it. Different GTAs in the same path may contain different elements, depending on measurement availability.

Location and time may be represented in the GTAs by many different methods, but the Coordinated Universal Time (UTC) and WGS84 location standards [6] are convenient choices that are commonly used by commercially-available devices. WGS84 describes location in terms of latitude and longitude on an oblate spheroid model of Earth, and specifies elevation as the distance above or below this model. UTC provides a time index that is independent of time zones, thereby simplifying comparison of times across different paths. To display "local time", one must first add the appropriate UTC offset (in hours), which may be determined from a database that maps locations to time zones. Capture device heading and tilt may be measured via an electronic compass and inclinometer, respectively, while field-of-view may be estimated from the camera focal length (zoom) and imager properties.

We store the media portion of PEM as individual files in standard formats such as MPEG, JPEG, and WAV. This allows easy construction of path-enhanced media on top of existing database and file directory formats, and easy integration of PEM with standard media players and software packages. If supported by the format standard, each media file may include internal tags such as subject keywords, creator identity, capture time and location, and access permissions. In some appplications, to allow for faster access and search of the set of media associated with a trip, we also store within the PEM a separate table with start times for and direct links to all the media.

We allow a given path to be divided into "segments", each composed of a time-ordered series of GTAs. Each segment has a "header" that can specify owner or creator identity, a descriptive title, access permissions, display style, or other properties. This hierarchical structuring of paths in terms of segments and GTAs better accomodates the sporadic nature of movement during extended trips, and allows paths recorded by different people or different devices to be compiled into a single logical path more easily.

We refer to one logical unit of path-enhanced media - built around a single, optionally segmented path, and often corresponding to a single trip - as a *PEM Object*. Figure 2 shows an XML representation, similar to that used by our prototype systems, of a PEM Object.

Each PEM Object may be associated with one or more *View* data structures. Different types of Views allow the same PEM Object to be organized for display and interaction in different ways. "Map Views", such as that shown in Figure 1, overlay representations of path and media on a suitable map stored within the View. We focus on Map Views in this paper, but other View types of interest include

```
<path>
  <location_coord_sys>WGS84</location_coord_sys>
  <time_coord_sys>UTC</time_coord_sys>
  <segment>
    <author>John Smith</author>
    <device>John's HP2200</device>
    <perms>744</perms>
    <gta><date>11/03/2001</date><time>22.14.00</time><lat>37.42510</lat>
         <lon>-122.17269</lon><elev>676.56</elev></gta>
    <gta><date>11/03/2001</date><time>22.14.06</time><lat>37.42512</lat>
         <lon>-122.17275</lon><elev>676.98</elev></gta>
    <gta><date>11/03/2001</date><time>22.14.12</time><lat>37.42513</lat>
         <lon>-122.17275</lon><elev>677.03</elev><heading>187.06</heading>
         <mediaFile>dsc0045187.mpg</mediaFile></gta>
    <gta><date>11/03/2001</date><time>22.14.18</time><lat>37.42515</lat>
         <lon>-122.17276</lon><elev>677.04</elev><heading>188.56</heading>
         <mediaFile>dsc0045187.mpg</mediaFile></gta>
    ...
  </segment>
  <segment>
  ...
  </segment>
</path>
```

**Fig. 2**. Example PEM input fragment in XML format. Capture of a movie begins at third path location. The proportion of GTAs without media is much higher in practice than is shown here.

ones displayed with a calendar-like format and others organized by media type (e.g. photo, video, or audio). A View may also be used select a subset of the PEM Object data to be displayed. For a PEM Object that collects data from several members of a vacationing family, one View may show only the path and media captured by one family member, while another View might sample from all members and show what the family judged the "highlights" of the trip to be. Views may thus decompose a trip into meaningful "layers" that can be added or removed from display during interactive browsing.

## 3. MEDIABEADS ARCHITECTURE

An overview of the MediaBeads architecture for working with path-enhanced media is shown in Figure 3. It includes components for capturing, viewing, editing, augmenting, and querying PEM data, which will be described in subsections that follow.

### 3.1. PEM Capture

Candidate devices for PEM capture include a video/audio-capable digital camera with on-board GPS receiver, and a cellular telephone with integrated camera and GPS. Such devices are available from Ricoh, Kyocera, Toshiba, and others, but currently do not record full path. Until an integrated PEM capture unit becomes available, path and media may be captured with separate devices - such as a GPS receiver and a video/audio-capable digital camera - and combined into a PEM structure via software. To properly align the two data streams, one must simply set the internal clocks of the two capture devices to the same time prior to use. Where the reliability of GPS data is poor, such as indoors or in deep canyons, it may be complemented with alternative location-determining technologies such as wireless beacon triangulation [7], analysis of cellphone signal profiles [13], or person tracking with video cameras [3]. In practice, we have found that path interpolation well covers many of the faults of GPS, and that details of indoor paths are often unimportant to users.

MediaBeads does not require, however, that the input PEM reflect a real, physically traversed path. Paths can also be constructed by allowing users to draw on a map, perhaps connecting indicated sites at which they captured media via their best recollection of the streets, trails, and other routes they traveled. In other contexts, such as when planning a trip or generating queries of PEM databases (see Section 3.4), people may wish to describe entirely fictitious paths to which media and other data may be attached.

### 3.2. PEM Browsing and Editing

For browsing and modifying path-enhanced media, a "Map View" of the data is often most useful. MediaBeads' "Static PEM Renderer" builds such views, as shown in Figure 1, in a series of steps:

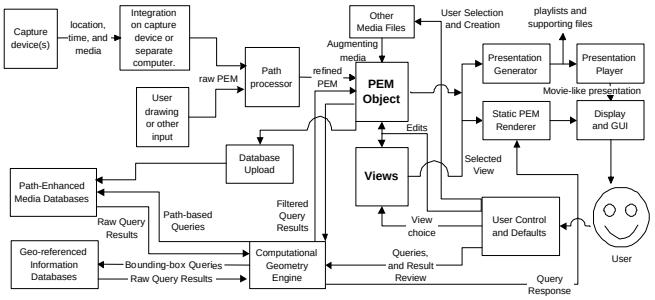


**Fig. 3**. MediaBeads architecture.

**Path processing:** A "Path Processor" module optionally smooths the path to remove spatial errors (e.g. due to multipath signal interference at the GPS receiver), fits it with splines or other curve descriptors, interpolates it within regions where measured location data is not available, and simplifies it to remove small loops or meanderings (perhaps where a person back-tracked briefly or otherwise dawdled). "Stop points" along the path, where little or no travel occurred for a designated time, may be detected for subsequent display as potential interest points along the trip.

**Map retrieval:** The bounding box latitude and longitude for the PEM path and media is calculated, and an online map server is queried to retrieve one or more appropriate maps that cover the box. Maps are tiled or blended at borders, as required. Street maps, aerial photographs, tourist guides, or other types of maps may be used, provided that they are scaled similarly to the WGS84 coordinate system in which PEM locations are specified.

**PEM overlay:** The path is overlaid on the maps as a prominent continuous curve, rather than as a series of location points. Each media file is represented with an icon placed near its capture location on the map, with an optional connecting line to the path location of its capture. For non-instantaneous media such as audio or video, the path interval over which the media is captured is highlighted. Media icons may indicate the media type (e.g. an image of a camera for photos), may include text indicating the capture place, time, or person, and (for visual media) may include image thumbnails or video keyframes of the media content itself.

**Field-of-view:** When information such as capture device orientation and focal length are available, the FOV of the capture device during visual media recording is estimated, and indicated on the map as an arrow or 2D shape extending from the capture location.

The resulting Map View of the PEM data may be scrolled or zoomed for examination, and individual media files may be selected for playing, if their permissions allow, by clicking on them with a mouse.

The Map View is also convenient for performing many editing operations of PEM data, including redrawing portions of the path, attaching new media files to the path, opening media files in external editors (e.g. for photo cropping or noise removal), changing the location or FOV associated with a media file, creating and editing Views of the PEM, "bookmarking" significant path locations and media, and modifying access permissions and ownership. In browsing the displayed PEM, a user may recall his thoughts at some stop point, or he might remember an interesting anecdote that is the cause of an unusual bend in the path. The editing interface allows the user to add these memories to the PEM, by creating new text or voice annotations and attaching them to path locations or intervals.

Map Views of PEM data also provide a powerful, versatile interface to spatiotemporally-indexed databases accessible via the Internet or through pre-packaged media such as CD-ROMs. A "Computational Geometry Engine" component of MediaBeads converts user queries based on path points, path intervals, or selected media into forms (typically in terms of spatiotemporal bounding boxes) understood by these databases. The Computational Geometry Engine also receives raw query results from the databases, and filters and formats them for graphical presentation on the Map View. The user may then browse the results, and optionally select some to be integrated into the PEM via automatic modification or addition of path GTAs. This might be used to augment a PEM with stock photos of what one was not allowed to photograph inside a museum, or to replace a picture of a landmark that did not turn out well due to lighting or weather. Similarly, one can search for and augment the PEM with names of restaurants visited, audio of regional music heard, or news headlines from the date and place of the trip. One may also just browse the query results, to search for currently available hotels, check today's weather, or see webcam views of what a place looks like right now.

### 3.3. Automatic Trip Presentation Generators

Perhaps the most compelling use of path data is for generating, with little user effort, trip presentations that may be enjoyed much like a television program. From the above-described Map View of PEM data, it is easy to envision how such a presentation might progress. Specifically, an icon representing a traveling person can move along the recorded path, with the media attached to the path being played when the icon reaches it. Media play can be concurrent with path movement, to provide more "action" in the presentation, and music or narration preferably accompanies it. Viewers of such a presentation have the feeling of "re-experiencing" the trip, and the sense of context is often much stronger than in conventional methods of media-sharing or trip-reporting.

MediaBeads contains a "Presentation Generator" that creates playlists of events such as animating a traveling icon, initiating play of a captured video, geometrically transforming (e.g. scrolling or zooming) a map-based view of the PEM data, or starting up a background music clip. The Presentation Generator output may be thus represented compactly as a collection of media files and descriptions of items to be rendered (e.g. path drawings, traveler icons, etc.), together with timing information, in a format akin to SMIL [11]. The MediaBeads "Presentation Player" reads this description to draw, animate, and play the presentation. The output of the Presentation Player - namely, a video and audio stream - may be fed to conventional media players, or it may be stored, for example on a DVD, for easy distribution and future playback.

The steps for generating the playlists include those for creating static Map Views of PEM data as in Section 3.2, plus others accounting for the dynamic nature of the presentations, including:

**Media clustering:** People typically capture media in bunches, in response to the occurrence of an interesting event or the arrival at a significant site. MediaBeads clusters media within PEM into groups that were captured at nearby locations and times, and explicitly indicates these clusters as icons on the map with names derived from location databases or user notes. During the presentation, zooming of the Map View toward a cluster icon causes it to disaggregate into individual path elements and media files.

**Path progress indicators:** As the presentation plays, travel progress along the path is indicated by either changing the color and/or thickness of the path, or by moving an icon along it. Icon choice may vary during the presentation to reflect the underlying terrain or travel speed: for example, an airplane icon for fast travel, or a "walking person" animation for slow travel in an urban area. In some applications, it is useful to move the icon at a rate proportional to which the path was actually traveled, while in others it is preferable for MediaBeads to automatically detect and shorten playback of "transit" or "sleep" segments with no media capture.

**Audio-visual overlap:** If audio recordings without video were captured, MediaBeads attempts to play the audio during display of photos or silent videos that were captured at nearby times.

**Transitions:** Visual and auditory transitions between different captured media, or between display of the map rather than of visual media, must be generated. In some presentation styles, the underlying Map View is alpha-blended with any playing media, such that the map (with a possibly moving path progress indicator) appears faintly beneath playing movies and photos, but returns to full strength when the media stops playing.

**Background audio selection:** Music is the most common choice for presentation soundtracks, but news audio clips - retrievable from geo-referenced databases for relevant locations and times - may also be appropriate. MediaBeads aligns background audio clip transitions with those between the visual elements of the presentation.

### 3.4. PEM Databases

Interesting possibilities arise upon gathering PEM data across many different trips by many people to many places. It is not obvious how to structure a database of PEM data for easy search, addition, and modification. We would like to support not just simple "bounding box" queries, but also queries that seek to compare paths or find data near selected portions of paths. The database should be indexed not just by location, but also by time. To achieve this, we divide space and time into a hierarchical oct-tree representation, where the 8 children of each tree node split the spatiotemporal domain of their parent with one division each in latitude, longitude, and time. In this way, given a point in space and time, we can quickly proceed from the root of the tree to the leaf whose spatial and temporal bounds encompass this point. In each leaf, we store a list of "PathIntervals", each of which is a link to the portion of some PEM path that falls within the spatiotemporal bounds of the leaf. The PathIntervals may be traversed to search for media, path, owner, or other data sought by the query. Queries requiring access to multiple tree leaves may be done efficiently with the appropriate sub-tree traversal. Queries based on path portions, rather than points, are decomposed into queries based on PathIntervals with divisions aligned with that of the oct-tree.

One useful application of such a database is to search for pictures of oneself taken by other people. Because a person's path is present in the database, we can look for photos at similar locations and times regardless of whether or not the person being searched for was recording media at those points. Field-of-view data present in the path can be used to further refine selection of candidate photos. Conversely, a person taking a photo of a crowd may ask, "Who is in my photo?", in which case path data from other people is compared to the location, time, and FOV of his photo. Such searches may lead people to wish to contact each other for chat or information, which may be achieved (if permitted) using PEM ownership information in the database. This may further lead people to search for others who took roughly the same trip at the same time as themselves, in order to compare stories or to find the identity of an interesting travel acquaintance. Finally, the database allows the possibility of complete "casual capture" of a trip, in which the user records only his path, and constructs a multimedia presentation of his trip from photos, audio, and video captured by other people at the same places and times.

## 4. EXPERIMENTAL PROTOTYPES

At present, we capture media and path with separate devices - namely, an HP video/audio-capable digital camera and a Garmin Etrek Summit [2] GPS receiver, respectively - and integrate the data into a PEM structure via software. The PEM is then fed as input to several prototype systems, collectively called *PathMarker*, that we have developed to implement many of the MediaBeads architectural components. Implementation and results for PathMarker are detailed in a companion paper [10]. In brief, PathMarker contains one system for generating animated presentations of trips overlaid on 2D street or satellite maps, and a second system for creating 3D "fly-bys" of media placed on billboards along a path in a 3D digital elevation model of the trip terrain. The former system also supports adding of voice and text narrations, and attachment or removal of media to the path. Experiments on many real, captured trips have yielded pleasing results for both systems, although much remains to be improved.

**Fig. 4**. Frames from a DVD presentation of a trip to Lake Tahoe, showing the transition from path traversal to viewing a still photo.

A presentation of a trip to Lake Tahoe, output by the 2D version of PathMarker, was used to create an interactive DVD. One may replay the DVD to see animations of travel along path portions, interspersed with media playback and transitions. Alternatively, the remote control can be used to randomly access different portions of the trip, and to browse or select media for viewing. A frame sequence from the DVD is shown in Figure 4. The DVD was found to provide an easy way to enjoy and remember a family adventure, and the map context was helpful in distinguishing between many photos of pines, mountains, water, and snow.

## 5. CONCLUSIONS

To the best of our knowledge, all prior work on spatiotemporal tagging of digital media has largely ignored the path information *between* points of media capture. We emphasize that some of the most interesting events happen when "the camera is not on", and we show that path data can greatly enrich the utility and enjoyment of digital media. The novel concept of "path-enhanced media", together with the MediaBeads architecture for manipulating it, helps people to view a collection of media as a unified "story" that may be augmented, presented, or related to other types of data in powerful new ways. We hope to collect PEM data from a greater variety of sources, and to develop more advanced systems embodying the architecture.

## 6. REFERENCES


[1] D. Diomidis. "Position-annotated photographs: a geotemporal web". *IEEE Pervasive Computing*, 2(2), 2003, pp. 72-79.

[2] Garmin Corporation, http://www.garmin.com.

[3] M. Harville. "Stereo person tracking with adaptive plan-view templates of height and occupancy statistics". *J. Image and Vision Comp. (22)*, No. 2, pp. 127-142, 2004.

[4] K. Hewagamage, M. Hirakawa. "Augmented album: situation-dependent system for a personal video/image collection". In *ICME'00*.

[5] H. Keranen, T. Rantakokko, Jani Mantyjarvi. "Sharing and presenting multimedia and context information within online communities using mobile terminals". In *ICME'03*.

[6] P. Longley, M. Goodchild, D. Maguire, D. Rhind. *Geographic Information Systems and Science.* John Wiley and Sons, 2001.

[7] N. Priyantha, A. Chakraborty, H. Balakrishnan. "The Cricket location-support system". In *MOBICOM'00*.

[8] Red Hen MediaMapper, http://www.mediamapper.com.

[9] P. Rigaux, M. Scholl, A. Voisard. *Spatial Databases with Application to GIS.* Morgan Kaufmann, 2002.

[10] R. Samadani, D. Mukherjee, U. Gargi, N. Chang, D. Tretter, M. Harville. "PathMarker: systems for capturing trips". In *ICME'04*.

[11] "Synchronized multimedia integration language (SMIL 2.0)". W3C Recommendation, August 2001.

[12] T. Smith, D. Andresen, et. al. "A digital library for geographically referenced materials". In *IEEE Computer*, 29(5), 1996, pp. 54-60.

[13] SnapTrack Inc., http://www.snaptrack.com.

[14] K. Toyama, R. Logan, A. Roseway, P. Anandan. "Geographic location tags on digital images". In *Proc. ACM Multimedia*, 2003.