

White paper

VES: An Environment for Visual Experiences

VES Group

Experiential Systems Laboratory
Department of Computer Science
University of California, Irvine
Irvine, CA 92697

November 2006

VES: An Environment for Visual Experiences

Objective: This white paper presents early ideas on concepts, functionality, and approaches to solve a problem that has become very important and is currently being addressed using approaches that have been at best partially successful in the past. Due to proliferation of digital cameras, including those called mobile phones, combined with easy environment to store and communicate these, people are collecting enormous volumes of photographs and videos. Unlike old ‘Kodak Moments’ remaining stored in shoeboxes, these ‘visual experiences’ keep growing on hard disks. Many approaches have emerged, both in industry and academia, for organizing and retrieving these photos and videos. These approaches range from feature based approaches to tags. The image and video search tools appear as two items right next to the ‘Web’ above the keyword box in all major search engines. That only helps in realizing how primitive current image and video (visual) search is. We decided to launch an ambitious project in this area by looking at the image and video search from a new perspective and bring together researchers to develop approaches to help in this.

This white paper outlines our perspective and approach to bring visual experiences at the same, or better, level as current textual techniques are. This is not a research paper. Our goal here is to present ideas, not completed work. The main objective of this paper is to share ideas among current team members of this project and provide an environment to discuss issues in a shared perspective. This white paper will be periodically upgraded by Ramesh Jain.

1 Introduction

The last few years have witnessed enormous growth in digital cameras and availability of storage devices. This has resulted in exponential growth of digital photographs and videos. Visual data, both photographs and videos, is rapidly becoming a primary medium along the side of text. Technology to acquire, store, and communicate visual data has

advanced faster than the technology to manage, search, share, and understand it has. Current search engines do respectable job with text data, but are pathetic with visual data. This project is to focus on developing technology for bringing visual experiences to the same level as text based information is. In the early phase we will only deal with photographs. The next stage will focus on videos also.

To cope with rapidly growing number of photographs, many album systems have been developed and deployed. Some commonly known systems are: Picasa, Corel, Adobe and ACDSsee. Many of these systems work as a photograph collections and provide basic image enhancing and processing tools. Recently a different class of photo systems has become popular. These systems, like Flickr, provide the environment to tag photos and effectively share them with others. Sharing of photos is a vital social activity and this has been popularized significantly by the availability of Flickr.com. The role of photographs in social networking and citizen journalism has been brought to the center by photo sharing systems. These systems use tags for searching of images. The popularity of tags in sharing and searching in these systems has been an interesting phenomenon. Use of people assigned tags is commonly called *Folksonomy*, as contrasted to taxonomy, and has been receiving attention in many circles.

Current photo album systems (Appendix A) are usually in two distinct categories: web based and local disk based. These systems provide the solution to upload pictures to a website, and organize the pictures in distinct collections. Each collection can be edited (add/delete/correct pictures), the captions can be added to any pictures for searching and sharing. People can organize pictures in many albums and present each album as a video or a slide show with background music and other features. These albums could be either manually prepared by drag-and-drop or by using logical selection based on tags or any other attributes.

For sharing online photographs, a user may invite friends or family members using e-mail/mobile phone and any communication tools to receive the information and pictures when they are added to the site. The recipients can access the pictures and view the web

pages and give their feedback and comments. Some sharing systems provide evaluation ratings. They assign the degree of ‘interestingness’ and this could be utilized for searching a class of pictures. Many interesting features are being introduced in this area. Flickr is a clear leader by a wide margin.

Local systems can help in organizing the pictures in the directories of local disk. These photographs can be edited, annotated and searched. One can find a picture and send it to his family and friends, burn it on a CD/VCD/DVD disc, and print the hardcopy or view the photos in TV using slideshow. To be actually able to efficiently search and find photos with these tools depends on tedious manual annotation efforts by the user.

There is a clear need for a system that will intuitively organize and manage digital photos, and share them with either a limited group or with everybody. This system will provide the functionality that allows it easy sharing of their photos with others and even put those photos in public if so desired. And all this should be effortlessly done in an environment in which the best practices of all existing photo systems are implemented. We are embarking on developing such a system by incorporating best concepts and techniques from the existing photo albums, photo sharing systems, image processing and understanding, databases, search engines, and social networks. Such a system may combine local as well as global storage of visual content as well as the metadata created by the owner and even by other people. Given the way technology and ownership concepts are evolving, there may be a need to explore concepts even in this area. In addition to building a complete system, we also adopt a perspective on photo management that is unique and in line with evolving concepts, and we believe it offers several long-term benefits.

Photos are taken to freeze a moment in time so one could experience it later. Remember the famous phrase popularized by the leader of the last generation of the photography technology – the Kodak Moment. We want to utilize this insight and build on that to bring out the best experiences of those moments to people effortlessly, efficiently, and in

most entertaining manner. In this project, we propose to use event based approach, therefore the first question that needs to be addressed is: what is an event?

1.1 Events and Objects

We believe that what people are interested in are storing and sharing their experiences. People experience events in which different objects participate and they want to store for later re-experience and for sharing it with others. Most experiences are related to events and objects intimately.

1.1.1 Events

The term 'event' has different meaning to different people. The meaning of event depends on the context and the granularity used in that context. There is inherent ambiguity associated with the term event as seen from its multiple usages. Wikipedia (<http://en.wikipedia.org/wiki/Event>) defines it as:

An **event** is something that takes place—an occurrence and arbitrary point in time. The term also refers to a significant occurrence or happening, or a social gathering or activity.

Dictionaries and encyclopedias are as confusing as the regular usage of the term is. Events are indeed different in different contexts. A wedding is an event, so is the wedding reception, and the cake-cutting. Also the first meeting of the bride and groom is as much an event as the birth of the bride, and the wedding of her parents, and so on. And yes the world cup Soccer final between Italy and France was an event and so is my grandson's first kick in his backyard. My pressing the key to type the next word is an event and so will be the posting of this article and then comes the event of your reading this and then reading this part and thinking that I am being silly! Theoretically even my moving finger to a specific key is an event. So events depend on context and take place at different granularities or resolutions.

Events are combined in many different ways to define event. And these combinations may again be combined with other events to define yet another set of events. So this

process of definition of events continues. These definitions are clearly determined by an application.

On the other side an event is the result of one or more past events, which were in turn results of other events and so on. Similarly, an event may result, maybe in combination with other events, in multiple events, which in turn may result in many other events. So this process of creation of events is a process that has been going on and will continue in future.

But, how do we define an event? What can be considered an event and what can not be? If all these things are events then how can we capture that in our computing systems – or can we? On first thought it appears that this is a confusing situation! This situation is not new, however. We are all familiar with Object Oriented programming and systems in computing. However, objects are equally confusing. Objects could be physical things or a concepts. Also objects could also exist at many resolutions. Once again Wikipedia (<http://en.wikipedia.org/wiki/Objects>) clearly shows how confusing it could be. So, I am an object, and so is the shirt I am wearing, and the buttons on the shirt. I will not get into more details of this – you get the idea.

1.1.2 Objects

We know that concept of objects have played a key role in the development of computer science. Object-oriented programming and object-oriented design concepts have dominated computer science for more than a decade. And this has been a powerful paradigm. Once again, Wikipedia gives a good definition of objects in the context of computer science:

Objects are language mechanism for binding data with methods that operate on that data.

See how nicely any reference to any physical or conceptual or any other kind of objects that we use in our regular language is avoided and an elegant new functional definition is

provided here. Objects become a mechanism for binding data with methods that operate on that data.

For computers, objects must be defined more precisely. In computer science, one defines objects as having two important components: data associated with the object and methods that operate or access data in predefined ways. The data is usually accessed only through the methods. In programming, one first defines classes of objects where all data associated with objects is defined, along with its type, and methods that will operate on all these data fields are defined. Each object in a program is an instance of a class. Each class may have subclasses and these subclasses may inherit some data types and methods from their parent class.

1.2 Defining Events

An event, defined in computing environments, should be a mechanism to define three important aspects of an event clearly and explicitly.

- Basic event data
- Information about the event;
- Experiences related to the event;
- Structural and Causal relationships with other events.

An event in computational form should represent data associated with the above aspects and processes to acquire and present these as may be needed. By providing flexible and expressive mechanisms to define these components and associated methods, one could define events effectively. The event environment should provide tools to define any event of interest from many disparate application domains. One may first define event classes and then each event in the system may be an instance of a class.

The basic characteristics of an event will be the ID of the event, its time and location. Time and location become fundamental defining characteristics of an event. A similar event may take place at different time and space and will be considered a different event. In this sense, an event is defined in spatio-temporal space. There are good reasons to

consider events as either point events or interval events. Point events are just points in the spatio-temporal (referred to as ST) space; interval events are regions in ST space. To make the terms that we use in this white paper clear, we provide the glossary in the Appendix B.

In this project, a photo-taking event is considered as an atomic event in the context of photo management. Clearly, photos are taken to capture experiential aspects of some objects or a scene. In fact, events and objects are very complementary and synergistic. This approach benefits the design and development of the system, the maintenance of the data and all associated information, and provides a unified approach to organization. Moreover, the biggest beneficiaries are the users who can search, share, access, and experience all photos in an experiential environment [5, 6].

The architecture of our system is illustrated in Fig.1. The technical details about how to implement this system will be addressed in section 4.

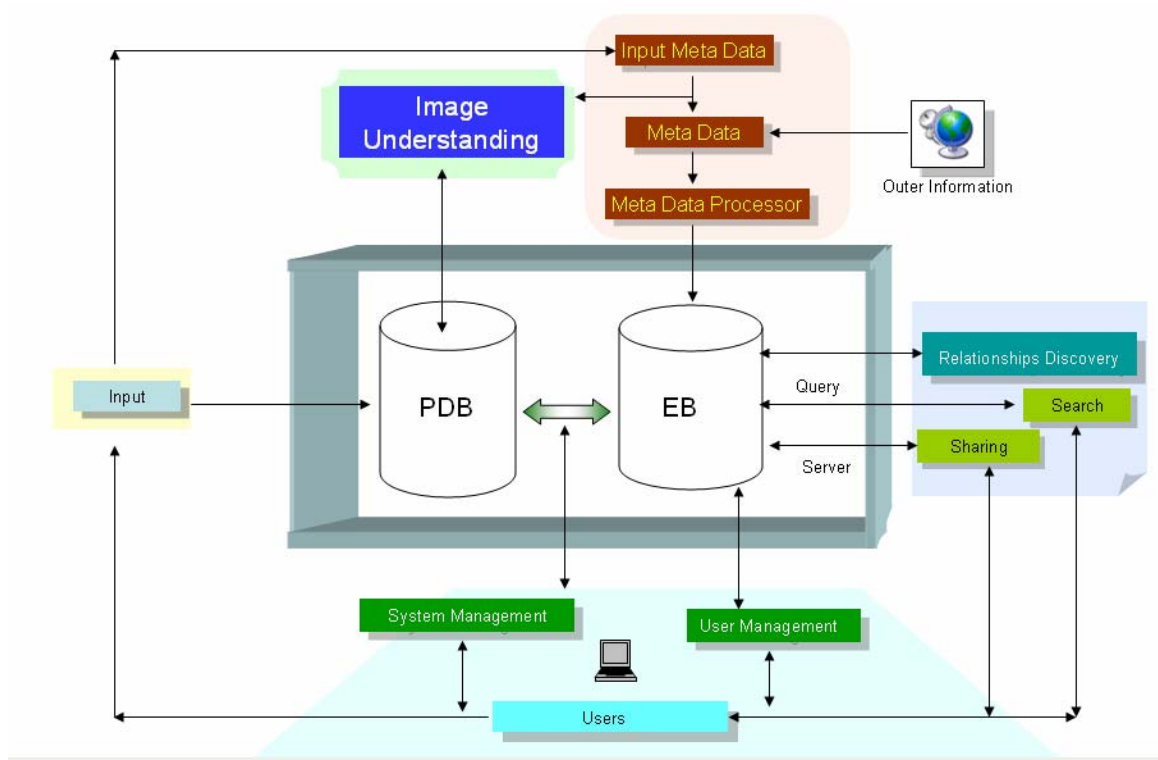


Fig. 1. Architecture of the system

As shown in the above figure, the event and picture databases, called Event Base (EB) and Photo Database (PDB) respectively, are at the center of the system. All data are stored in them. These data are entered through different mechanisms. Some event data are created through different combinations of metadata processing, image understanding, or manual annotations and tagging. Event mining techniques work on the eventbase periodically to update existing events considering new events in the system and to create any new high-level events as may be required. This process of creating and updating events is central to the operation of the system. A combination of search and sharing environment is provided for interaction with the system. The requests from a user are promptly tackled so as to offer good quality presentations based on personalized settings. The goal of this system is not to provide just a photo or a group of photos to a user, but to provide a complete experience of events and related objects to people.

In the first version we are restricting our experiential data to photos. Once we have a good understanding of issues related to developing such systems, we will include audio and video in the system. Later other kinds of experiential data will be included. At this stage we will restrict our system to only photos and focus on organization, management, and presentation of experiential data in context of events.

2. Organization of Photos and associated information

As mentioned above, the core of our system is event-centric management of information. Photographs are part of the experiential component of event information. This perspective allows us to associate all other information with photos in a very natural and unified manner. This also helps in organizing, accessing, sharing, and presenting photos with their contexts in more flexible ways.

We adopt the event model defined by Westermann and Jain [45,46]. This model is shown in Fig.2. and is used to build the EventBase and collect all photographs and associated information.

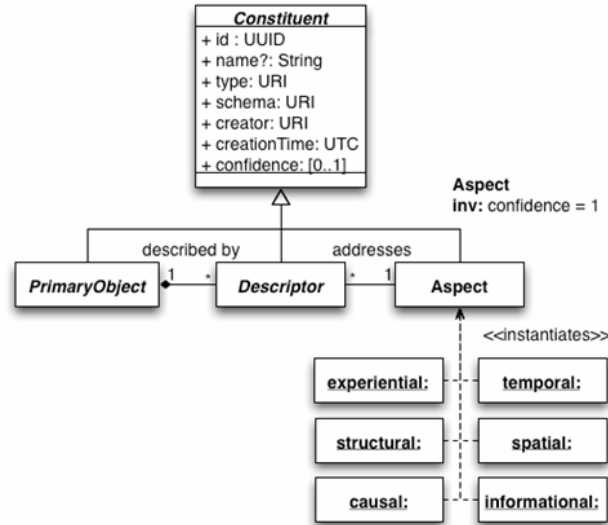


Fig.2. Event to organize photos

- **Basic Characteristics**
 - Event ID
 - Time
 - Location
- **Attributes**
 - Informational aspects
 - Experiential aspects
- **Relationships**
 - Structural relationships
 - Causal relationships

There are several major differences in representing events using the above structure for storing all data associated with events. A major advantage is that this structure removes focus from media such as photos, videos, or audio and organizes information and experiences around events. Thus an image event, a video event or an audio event is just media events and is not important in the real world. Those media events may be detected only to determine the event in the real world. By removing this focus from individual media and focusing on event, we can use more powerful techniques to correlate

multimodal data.

The above event representation allows capturing basic attributes of an event and allows building relationship structures among events that provide a multitude of flexible organizations to present events in many different formats. This structure permits us to group events in more flexible organizations than the traditional album structure. Albums in photo systems take the form of a folder in which each photo is placed. Normally a photo belongs to an album, but efforts are underway to allow flexibility there. In the new structure, people can have many flexible constellations of photos and present them depending on the principles employed in preparing the constellations.

In our system, a photo is represented as an experiential attribute of a photo-taking event. Some other characteristics of this event are easily available from metadata such as EXIF, some will be soon available, and other attributes may be computed using different techniques. Image characteristics can be computed using appropriate image understanding techniques, while other information can also be associated using different manual, semi-automatic, or full automatic techniques. All photos and associated data with photos and photo-taking process are used to populate an EventBase.

This EventBase can have many processes working on it to fill in all basic characteristics and attributes. The relational characteristics should also be filled in using different types of processes. All this information is maintained in the EventBase. One of the most important processing working on this will be analyzing events that come in the EB and grouping them into new events. Thus many photos taken in a wedding could be grouped into an event – *wedding ceremony*; while others may be part of the *wedding reception*; and yet others *wedding dance*. All these wedding events on that day could be grouped into the event called *wedding day*. For those who went to attend this wedding to San Diego, they may further have an event *Suzi's Wedding* which may have pictures not only from the wedding day but the rehearsal and other sightseeing events. All these grouping of events could be using different event models that represent composite events in terms of their constituent events. Obviously, an atomic event is the one that does not have any

constituent events. Composite events are created by considering structural and causal combinations that are meaningful in a given context and for specific people. For each composite event, separate event representations using the approach mentioned above will be used. And all such event will be created and added in the system. Thus, an EB is different from a database in the sense that it is not only for storing persistent data. The EB is organic and grows as new events are added. Also, in the long term, as new events are added, some old events may be modified.

The EB allows creation of multiple constellations and their presentation using appropriate presentation mechanisms. A user may define a particular event or present multiple events in which some participants or objects are common. This requires rich environments to interact with the EB. A user may also search or browse this EB using different types of tools.

Each event or each photograph may also be assigned an access code that will allow other people to access it based on their privileges related to it. Similarly, the owner of an event may decide to share it with others either proactively or passively, depending on relative inter-people privileges.

In this organization, the EB becomes the central organization. There are processes working to populate it and maintain it. On the other side, the system provides interaction environment to access and utilize the information in EB as needed.

There are several major advantages in representing events using the above structure for storing the events. A major advantage is that this structure is independent from the media, it organizes information and experiences around events. This structure allows us to group events in more flexible organizations than that of the traditional album systems. As an example, current photo albums take the form of a folder in which photos are placed. In the new structure, people have lots of flexible constellations of photos for storage and present them using different contexts defined in event and related characteristics.

All event information in this system will be maintained in the EventBase. One of the most important functionality is to analyze those events that come in the EB and compose them into new events. Event grouping could be using different event models that represent composite events in terms of their constituent. Obviously, an atomic event is the one that does not have any constituent events and will be marked so. For each composite event, separate event representations depending on the type of event will be provided. And all these events will be created and added in the system. Thus, an EB is organic and new events as detected from the growing data set are added.

3. Functionality

Our objective in this project is to include all best practices from existing systems and develop an environment that will be a powerful visual experience management system. With this in mind, the initial version of the system will have the following functionality:

- Import photos from
 - Cameras
 - Mobile phones
 - Computers
 - Scanners
 - Web sites
- Add descriptions to each event:
 - Tags of image regions (detecting ‘visual words’)
 - Automatic image annotations (detecting ‘visual concepts’)
 - Manual tags and opinions
 - Tags to complete images
 - Caption for each photograph
 - Fill in other attributes for photos from all sources of metadata
- Updating and creating events in EB
 - Detect composite events based on new events.
 - Creating composite events in EB and associating all properties

- Update related and causal event relationships
 - Update reports
- Creating and presenting collections by
 - An event
 - Topics
 - Themes
 - Relations among events
 - Event attributes
 - Visual attributes
 - Other social event aspects
 - Results of event exploration
 - Event attributes
 - Merging collections from different sources
- Photo search based on:
 - Key words or concepts of an event
 - Event elements
 - People
 - Location
 - Time
 - Event explorations
 - Feelings
 - Interestingness
 - Popularity
- Sharing photos
 - Sending photos with:
 - PPT, PDF Slides, Flash, SMIL, or SVG.
 - Blog/e-mail/MMS/Cell Phone/
 - Event based photo collection
 - Event presentation based story-telling
 - Monitoring availability from:
 - Send e-mails to friends

- Viewer's comments
 - On topics
 - Of specific types
 - Recommendations
 - To friends
 - From friends
 - Based on use history by known friends
 - Based on a person's prior use
- Management and editing
 - Fixing and enhance a photo
 - Adding/deleting events
 - Editing a composite event by adding, deleting, or modifying atomic event attributes and photos
 - Backup, maintenance, modification, merge
 - Sorting events in different criteria
 - Drag and drop
- Presentations
 - DVD files using events with photo presentation
 - Web based photo collection
 - Wireless access (MMS) for mobile phones
 - Hardcopy
 - Multimedia presentations, SMIL, Flash, SVG, ...

The above list is likely to grow as new functions are requested by different class of users. The current list is prepared based on the functionality observed in current systems.

4. **EventBase Schema**

We have discussed the general event model and the functionality expected in VES above. In this section we discuss the event schema use in VES by considering different

information attributes, their sources and confidence (or trust) in sources and requirements. We discuss several important characteristics that should be part of the event model. The detailed schema using the model E is being designed.

A. **Photo ID:** Each digital photograph has a unique ID associated with the file. This ID will be used in all references to the photo in the system.

B. **Photo name:** Photo name will be the one assigned by the owner of the photo and may be used in references by the user.

C. **Photo path:** This item subsumes file name, extension and the URL of the photo. If a photograph is deleted/removed from the database, this item will be changed correspondingly.

D. **Camera Metadata:** Digital cameras save JPEG (.jpg) files with EXIF (Exchangeable Image File) data. Camera settings and scene information are recorded by the camera into the image file. This information may be stored in its original format with the photo in the photo storage. Appropriate information from EXIF may be imported by the EB at the photo entry time. The system should accommodate other types of acquisition time metadata.

E. **Time Info:** This item provides the history of creation/editing/uploading/last view time of a photograph. An important issue to be explored is also the fact that ‘viewing’ an event is also an event related to the original and should be considered in some contexts.

F. **Temporal Tag:** This item records the time relevant strings, such as season, festival and temporal activities.

G. **GPS Info:** This item gives the precise location values: longitude, latitude, and altitude.

H. **Address Info:** This information gives symbolic location description (building, street,

city, county, state, country) such as the address where the picture is taken.

I. Objects: Objects in a picture are important description of the picture and should be stored. Particularly people are a very important object in pictures. Number and names of people should be recorded. Our long term goal is to exhaustively record all possible objects in an image.

J. Visual Concepts: A picture may be identified as a concrete concept like 'city scene' or 'beach scene' or an abstract concept like 'peace' or 'love'. All such concepts in the context of the picture should be recorded in the EB.

K. Tags: Each picture may be assigned manual tags. A list of all such tags along with the identity of the person assigning tags should be stored. Additionally there may be automatic techniques used to assign tags. Automatically assigned tags should be so represented in the EB.

L. Comments: All comments and annotations should also be stored. Any descriptions such as popularity can be stored in this text item.

M. Image Features: Image characteristics that are considered relevant that may be required will be stored under this. These features may be either low level characteristics such as color histograms, or characteristics of major segments in an image.

N. Object Features: Characteristics of all objects may also be stored. Such characteristics may be color, texture, size, and shape of the object. In some cases, relationships among objects may also be stored.

O. Event Relationships: Relationship of this event to other events should be represented using appropriate characteristics. All subset and superset relationships (like in tree structures) may be represented using appropriate mechanisms. Causality of events should be separately represented from structural relationships.

P. Experiential Components: Each event has multiple experiential components. Pictures are just one. If there are other experiential components related to the event, they should also be represented using appropriate links and storage structures. Attributes of each such component should be stored under individual components.

Q. Derived Events: By event discovery or other methods higher level events may be defined. Such events should be appropriately represented. Note that such events may need slightly different fields. For example, such events may have a collection of photograph attached as an experiential component rather than just one photograph. Such events may also have a complete subgroup, or a sub-tree of events under it. We may need to define properties like inheritance and generalization for event attributes.

5. Technical Requirements

In this section we discuss technical requirements and challenges for our system. At this stage we are identifying different technical components. We identify major technical components and discuss what may go in these components. The details of the technology will evolve as we start working on the implementation of VES.

5.1 Ingesting Photos and associated metadata

Photos may come from various sources such as cameras, e-mails, web, and scanners. In each case the size of the image, the quality, and associated metadata may be very different. Techniques should be developed to extract whatever metadata is available to fill in the associated fields in the EB. In some cases, it may require use of special processes to interpret metadata from the source to be included in the event information.

The originals of all images should be stored in a photo storage system. A thumbnail, link to the original and other event related data should go to the EB. If the original image is

enhanced or processed, the new image will be stored as a processed version and will be entered as a new event. Only owner will be allowed to make these decisions.

If a user copies a photo, the system may offer to copy all event information related to the photo also. All the information will be stored using an Event Markup Language (EML). We are currently in the process of developing EML that will help converting event model used in an application to be represented in human readable and transformable representation.

5.2 Metadata Processor (MDP)

Some metadata come with an image from the camera or other sources. Some other metadata may be obtained by prompting the user to provide tags and descriptions. The names of the files could also be used as a form of metadata. This metadata will be used to fill in appropriate fields of the event model.

The task of the MDP is to explore all other sources of data that could be used to fill in other fields in the event model. Some common sources to use for this processing could be calendar and contact lists of the user, GPS based information about the location combined with maps, and other sources publicly available on the Web. In some cases our VES may itself have sources that could be used. Effectively, this may require a sophisticated network and architecture to implement rules for each source and also the information flow used in the system.

We expect to utilize approaches similar to those used in databases for defining constraints on different attributes, rules and functions to extract values of fields in event model using different input sources. Typically we expect that such tools will be defined considering different types of input media and different media resources. In the current version, however, we may restrict our media sources to photos and text only.

5.3 Detecting Visual Words and Concepts

Computer vision and pattern recognition techniques could be used to detect some Visual Words (basic objects) and Visual Concepts. Such techniques are advancing and if about 500 words and 200 concepts could be automatically detected from images, then a powerful automatic annotation and tagging environment will become available. Our initial target will be to develop a vocabulary of 100 visual words and 30 visual concepts that could be automatically detected with reasonable accuracy. We are starting an effort to define these words and concepts in a general setting and then extend them further.

Most current techniques for visual annotations do not use any information from the metadata. In fact, most such techniques have traditionally relied more or less completely on only the visual attributes available in a single photograph without utilizing any other information related to the environment in which a picture was taken or camera parameters. We believe that the metadata available from EXIF and similar standards and manual tagging and descriptions could be a very powerful source of information that could help in automatic visual detection significantly. More precisely, such information may initially help in selecting appropriate parameters in segmentation and then guide object detection processes. These approaches open up a very interesting research area for developing powerful context guided segmentation and object recognition techniques that use context parameters to select more effective and targeted processes.

5.4 Relationship Building

Events follow network characteristics; the value of an event increases exponentially in relation to the total number of events in the network. An event is like a node and its value increases exponentially as other events and relationships among them are added. Our event model allows relationships of different kinds, including causal. An important task for the system is to develop approaches that will allow discovery of these relationships among events. This is a very important research topic. We have already started doing some research in this area using Attribute-Space-Time clustering approaches. These

approaches will allow us to build some basic grouping of atomic events. These groupings can then be refined and further built using tags and other sources of information.

Causal relationships discovery may require stronger domain models. We will study roles of different models in discovering different types of relationships. Initially we may use models created manually and then only explore techniques to build these models using machine learning techniques.

We expect that independent relationship discovery processes will work on the EB and update such relationships. The facility of probabilistic representation for different attributes may be very useful in such relationship discovery approaches and for the representation of the strengths of the results of these processes.

Community effects, social learning, and emergent semantics:

The building of relationships of events can also be based on social learning. For example, a photo-taken event A may have a GPS position. An event that is related in topic and time may only have a position derived from a cell phone. So the latter photo-taken event might adopt the GPS position from photo A. Traces of events in space and time from different users can be grouped together and may also use the collaboration to refine the knowledge about the single (lower-level) events. Also the events could learn from my previous events, can learn from other user's events.

5.5 Search and Browsing Environment

VES will provide a combination of browsing and search environment for people to find photos they are interested in. We believe that this combination in event-centric environment will allow people to interact with the system more naturally. They will be able to articulate their queries in many different ways using events as well as pictorial characteristics. We are confident that for the interaction with large photo collections,

ideas of emergent semantics will be effective in such an environment. We intend to provide tools for such interactions.

Moreover, this environment will combine ideas from taxonomy, folksonomy, facets, and search based on events, words, and pictorial attributes and their combinations. We will also explore novel mechanisms for search based on compositions using visual words, basic image attributes and tags. We believe that current tools for interaction with photo collections are not effective and hence we are interested in exploring and finding such tools. This may require combination of search, databases, visualization and human interaction to converge on satisfactory approaches for design and development of such interaction environments.

Using the interactions of people, the system will develop queries to be processed by the system to return appropriate events as results.

5.6 Experience Presentation

The results returned by the system in response to queries generated by the interaction environment, system will create and present appropriate presentations. These presentations may be a photo with information and attributes, collection of photos and attributes in a specific way, a slide presentation with appropriate backgrounds and any such multimedia presentations created using photos, event information and other presentation material. Such presentations may also consider the context of the presentation in their automatic authoring environments.

The system may also consider the nature of the device on which the presentations may be made and the mode in which a user may want to receive the experience.

5.7 Personal Event Management

A person entering photos in the system may want to manage photos and other data associated with such events. This may include enhancing photos, entering tags and descriptions, entering relationships, adding locations, and any other event related information.

The system will provide easy environment to specify different access-level for events and define groups to share and access photos.

In addition to the above basic characteristics, tools should be provided to manage the access and sharing properties of the events, photos, and soon videos.

5.8 System Administration

The system should manage all above functions. It should also manage all the actions resulting from the sharing requirements. We anticipate that this functionality will use mostly current tools and techniques and hence may not require any significant research efforts from our side.

One area that we may want to consider seriously in this effort may be research in privacy and security. It may be important to develop tools so that privacy of people is not compromised. Our thoughts in this area at this stage are just to be sensitive to this issue and identify researchers who may help in identifying and implementing approaches in this area.

6. Creating and Accessing Event Experiences

A major benefit of the event-centric approach is organization of photos entered by different people at different places and at different times in the system in a unified way by assimilating all these around common events. The second major benefit to users is a very flexible and natural environment to access photos using many approaches. A lot of

details will emerge as the system is implemented and tested. Here we mention some typical example scenarios to start with.

Suppose there is a major conference event – MM 2006. People from 123 countries attend this conference and they all take several photographs at the conference. Since these people are aware of VES and are users of the system, they all enter these photographs in the system. They all use a common tag assigned to the conference, MM2006, when entering their photos. Obviously most people are entering these photos after going back to their homes and are doing that at different times. But since the system knows the tags, times, and location of all these photos, it can assimilate all these photos, though maintaining ownerships by their rightful owners, around one event. User Anon Y Mous may have taken only 5 photos at the conference but can enjoy total of 3456 different photos entered by different users. Moreover these photos could be organized in many themes using other metadata, manual tags, and similarity of visual words and concepts.

As another example, if a user wants to know how extensively Deepavali celebrated in Singapore, all that this user will have to browse or search for Diwali (change in name intended) in Singapore and the system will show all the pictures in the system and provide the user a good experience of the celebrations in Little India as well as in other personal celebrations (declared public by owners).

As just one more example, user Regu Lar has photographs for the last 30 years in the system. For his daughter's 25th birthday, she wants to make a presentation in the party showing the cake cutting on all her birthdays to show how she, and things, have changed over the last 25 years. All that she has to do is to just articulate this in a simple query and the system pulls out all those pictures for the presentation. Of course, there is a popular demand to show her pictures of blowing candles in all these events – and this is done as easily by presenting another query to the system. Now Mrs Regu Lar, being a true global citizen, has lived in 11 different countries and 16 different cities during those 25 years and has used about 30 different cameras. But all that does not matter – the presentations are made directly on their latest 105-inch plasma panel.

7. Current Status

We believe that time is right to build this event-centric system for photo management to provide a unified environment in which people can manage their photos. Our goal in this white paper was to present our current thoughts on different functions that this system will provide and technical approaches used to achieve this functionality.

We believe that this effort requires expertise and research talents from diverse scientific disciplines. One goal of this white paper is to bring together such people by defining and specifying the requirements and then connecting with right people. Our approach in this project will be inclusionary rather than exclusionary. We will encourage researchers to identify a problem, develop right approaches and test them in the context of this system.

This draft will be revised periodically and will be kept updated on our website. We will create a VESWiki to invite people to contribute for the continuous development and refinement of ideas related to this document. The document will, however, be maintained separately to articulate vision, uniformity and integrity of ideas, architecture, and the direction. The document will reflect team's thinking and ideas and will be updated periodically. The project will be coordinated by a person responsible for maintaining environment to build appropriate environment and infrastructure. All participants will be encouraged to contribute to the environment building.

8. The Team

Initially, Ramesh Jain will be the director of the project and Wei-Qi Yan will help him in coordinating it. The initial team (tentatively) consists of the research teams of the following universities and professors:

1. University of California, Irvine
 - a. Ramesh Jain

- b. Sharad Mehrotra
 - c. Max Welling
 - d. Padhraic Smythe
- 2. National University of Singapore
 - a. Tat-Seng Chua
 - b. Mohan Kankanhalli
- 3. University of Oldenburg, Germany
 - a. Susanne Boll
- 4. University of North Carolina, Charlotte
 - a. JianPing Fan
- 5. Rensselaer Polytechnic Institute
 - a. Sibel Adali
- 6. University of California, San Diego
 - a. Amarnath Gupta
- 7. University of Amsterdam, CWI
 - a. Arjen deVries
 - b. Marcel Reinders