Space–Time Environmental Image Information for Scene Understanding

by Arnold Tunick
NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer’s or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.
Space–Time Environmental Image Information for Scene Understanding

by Arnold Tunick

Computational and Information Sciences Directorate, ARL

Approved for public release; distribution unlimited.
The interpretation of space- and time-varying environmental image information can pose serious challenges for rapid and robust scene understanding, particularly for problems of interest to the Army. Changing environmental conditions, such as illumination, precipitation, and vegetation, can obscure features, degrade object recognition, and modify saliency and image context. Current methods for scene understanding do not address elements of image information (and image context) affected by space- and time-changing environmental conditions, and as a result, important and meaningful features of the image data may be overlooked. In this report, we propose that it is important to incorporate space- and time-varying environmental image information from the very beginning of the data collection process so that the recorded images can be more effectively indexed and retrieved for operational use and analysis. This top-down approach not only provides a systematic characterization of the measured data for better scene description, but can also help the end user (Soldier) develop improved course of action strategies based on scene understanding (algorithms and analysis) incorporating battlefield environments changing in space and time.

subject terms
computer vision, image resolution, context, saliency, visibility, illumination, vegetation, visual motion
List of Figures

Fig. 1 Primary space ($s$) and time ($t$) scales .....................................................3
Fig. 2 Can you correctly identify these images? Image resolution: a) 30 x 20 pixels, b) 30 x 14 pixels, and c) 30 x 16 pixels.............................................4
Fig. 3 Higher-resolution image scenes corresponding to the 3 shapes shown in Fig. 2. a) Taj Mahal (photo courtesy of desktopdress.com), b) US Capitol Dome (photo courtesy of Library of Congress), and c) nuclear power plant, Bushehr, Iran (photo courtesy of Behrouz Mehri/AFP/Getty Images). .................................................................4
Fig. 4 Can you correctly identify these objects? ..........................................................5
Fig. 5 Low-resolution images of the scene from which the objects in Fig. 4 were taken, where neither large nor small objects are discernible (left: 16 x 10 image pixels, right: 32 x 20 image pixels).................................................5
Fig. 6 Same images as shown in Fig. 5 but with a slightly higher resolution (left: 64 x 40 image pixels, right: 128 x 80 image pixels)...........................5
Fig. 7 Same images as shown in Figs. 5 and 6 but with an even higher resolution (left: 525 x 336 image pixel, right: 3888 x 2492 image pixels). Note that the hoist, surveyor, and engineers wearing hard hats in the far-field of the imaged scene all required increased image resolution to be clearly identified. .........................................................6

List of Tables

Table 1 Image resolution information (in numbers of pixels) ..........................6
Table 2 Time- and space-varying elements of scene understanding .................8
Table 3 Scene description indexing (top-level view)........................................9
Table 4 Available/accessible environmental image information..................9
Acknowledgments

I thank RE Meyers, P David, and G Warnell for helpful discussions. This research was supported by the US Army Research Laboratory.
1. Introduction

Rapid and robust scene understanding is a critically important goal for the development of Army autonomous intelligent systems.\(^1\) For outdoor natural scenes, it will be important for autonomous intelligent systems to be able to quickly discern the depth of view, navigability, exposure or concealment (as it relates to object searching), and transience, that is, the rate at which elements of the scene or its environment are changing in space and time.\(^2,3\) In this regard, saliency estimation has been helpful to computationally identify elements in a scene that immediately capture the visual attention of an observer.\(^4,5\) Several recent papers have discussed concepts associated with visual saliency to enhance automated navigation and scene exploration.\(^6-8\) Note, however, that the most active or salient object(s) in a scene may not represent the most important or meaningful feature(s) of the scene.\(^9\) For example, an automated vision system may readily detect changes in the ground surface as a new or different object in the field of view; however, recognizing the physical characteristics of the new surface\(^10,11\) (e.g., shallow or deep water, thick or thin ice, snow, mud, quicksand, etc.) and observing any changes in the environmental context of the image\(^12-14\) may be critically important. Characterizing interactions between objects and the environment also can contribute to physical scene understanding.\(^15,16\) In the example above, if a scene depicts vehicles or personnel activity in a changing complex environment, then robust scene understanding could provide key border and accessibility information for navigation and trafficability.

Nevertheless, many current methods for scene understanding, like those that generate image descriptions via automated semantic labeling\(^17\) or visual scene classification,\(^18\) do not address image information (and image context) affected by changing environmental conditions. Yet, the interpretation of changing environmental conditions can pose serious challenges for computer vision processes, such as those associated with place recognition, navigation, road/terrain detection, and scene exploration.\(^19-24\) This is because rain, snow, and fog weather events, smoke, haze, or other changes in lighting and visibility can significantly obscure features, degrade object recognition, and modify the saliency and image context of an outdoor scene.\(^25-32\) Naturally, scene-depicted environmental conditions can vary with time of day, season, and location.\(^33\)

Similar challenges can also extend to interpreting space- and time-changing scenes due to visual motion of objects within the field of view.\(^34-36\) In this case, changing environmental conditions such as illumination, precipitation, and vegetation can make feature recognition of moving objects unclear, so that identifying moving objects in outdoor environments becomes more difficult for vision-based intelligent
systems. As an example, poor contrast in images can be brought about by low visibility due to environmental effects or weak illumination, such as during dawn, dusk, or night. Visually degraded or blurred images can be brought about by rapid movements of the camera and/or objects in the field of view, especially in the low-light case, which necessitate longer camera exposure times. In addition, space and time variations in scene illumination can affect the optical flow field in images and movies. Note that there have been many optical flow approaches used to detect the motion of objects in a scene, which have been helpful in a variety of applications. Nevertheless, camera motion may introduce some unmanageable artifacts with some of these gradient-based optical flow approaches if they are not augmented by more sophisticated spatiotemporal analyses.

In this report, we propose that it is important to incorporate space- and time-varying environmental image information from the very beginning of the data collection process so that the recorded images can be more effectively indexed and retrieved for operational use and analysis. This top-down approach not only provides a systematic characterization of the measured data for better scene description, but can help the end user (Soldier) develop improved course of action strategies based on scene understanding (algorithms and analysis) incorporating battlefield environments changing in space and time. Incorporating space- and time-varying environmental image information for better scene understanding can be vital to support numerous Army missions, such as those related to weather elements on the battlefield that can alter terrain features and trafficability; low visibility that can impede reconnaissance and target acquisition or alternately conceal friendly forces maneuvers and activities; and wind speed and direction that can favor upwind forces in nuclear, biological, and chemical (NBC) attacks or decrease the effectiveness of downwind forces due blowing dust, smoke, sand, rain, or snow. In addition, reporting wind speed and direction information at the time images are being recorded can significantly influence the success of aviation-related missions, like those associated with unmanned aerial vehicle take off, landing, and in-flight control.

2. Space and Time Scales

Based on an analysis by Meyers, this section provides a framework to help categorize the spatial and temporal properties of recorded image data. Relevant time scales include, but are not limited to, the shutter exposure time, time interval between frames, time over which images are captured in a sequence, and time over which there is visual motion of objects inside the field of view. Space scales include, but are not limited to, the field of view, depth of view, image resolution, pixel size, pixel separation, scene color or shading variations as a function of spatial
location, spatial smearing of moving elements in the field of view, spatial smearing due to optical turbulence and environmental/weather effects, and smearing of textures in the field of view. Naturally, the smearing of elements in the field of view can also be related to the temporal resolution of the image data. Figure 1 illustrates the primary space and time scales that can be used to describe the various spatial and temporal resolutions of objects and/or activities in a recorded image scene, to include images recorded in varying environmental conditions. Here, $\Delta s$ and $\Delta t$ represent changes in position and time, respectively.

![Fig. 1 Primary space (s) and time (t) scales](image)

### 3. Image Resolution, Image Context, and Identifying Objects

To start, let us explore the impact of varying image resolution and image context on the analysis of an outdoor scene. Try to identify the 3 dome shapes in Fig. 2. Without some additional information related to the object size, texture, or shape or knowledge of how the image context may change in an expanded field of view, it is difficult to correctly identify and label these familiar images. Furthermore, distinguishing various image details, even in ideal conditions with regard to lighting and visibility, can depend on the image contrast and resolution, where image resolution here refers to the number of pixels that compose the image data input. Interestingly, Torralba\textsuperscript{51} reported that for human vision the brain can comprehend the gist of an image scene remarkably quickly, regardless of whether low- or high-resolution images are used. He concluded that images at the resolution of 32 x 32 color pixels can provide an observer enough information to correctly identify the semantic category and general layout of an indoor/outdoor scene. For example, in
Fig. 2 the main “dome” category for these low-resolution images is identifiable. However, if we consider Fig. 3, which contains expanded fields of view and higher resolution images from which the elements in Fig. 2 were taken, then the building domes and many additional image details can be identified over a much wider range of spatial scales.

Fig. 2 Can you correctly identify these images? Image resolution: a) 30 x 20 pixels, b) 30 x 14 pixels, and c) 30 x 16 pixels.

Fig. 3 Higher-resolution image scenes corresponding to the 3 shapes shown in Fig. 2. a) Taj Mahal (photo courtesy of desktopdress.com), b) US Capitol Dome (photo courtesy of Library of Congress), and c) nuclear power plant, Bushehr, Iran (photo courtesy of Behrouz Mehri/AFP/Getty Images).

To demonstrate this point further, Can you identify the 2 similarly shaped objects shown in Fig. 4 without some additional context? What if we look at the complete image (Fig. 5) from which the objects were taken? In this case, at low resolution, it is quite difficult to discern any individual elements in the field of view. Of course, the degree of image resolution needed for a particular task depends on the analysis or computer vision problem of interest. Yet, with regard to scene understanding and semantic labeling, the slightly higher resolution images shown in Fig. 6 clearly provide more usable information. In other words, when the image resolution is increased to 64 x 40 pixels and greater, one can more easily identify the layout and main elements of the image scene, such as the reactor dome and hard hat shown above. However, if still higher resolution images of this reactor site are considered (Fig. 7), then additional details and information may be gained, for example, intelligence relating to its operational status. By analyzing the extracted and labeled objects shown in Fig. 7, one might ask if the reactor site is still under
construction or near completion as evidenced by the engineers wearing hard hats, the surveyor, the hoist, and the electrical hazard sign. Note here that the hoist, surveyor, and engineers wearing hard hats in the far-field of the imaged scene all required increased resolution (i.e., $\geq 32 \times 32$ pixels) to be clearly identified (visually compare right vs. left in Fig. 7). Table 1 provides the various image resolution details (in numbers of pixels) for the labeled objects in this outdoor scene.

![Image of labeled objects](image)

**Fig. 4** Can you correctly identify these objects?

![Low-resolution images](image)

**Fig. 5** Low-resolution images of the scene from which the objects in Fig. 4 were taken, where neither large nor small objects are discernible (left: $16 \times 10$ image pixels, right: $32 \times 20$ image pixels)

![Higher-resolution images](image)

**Fig. 6** Same images as shown in Fig. 5 but with a slightly higher resolution (left: $64 \times 40$ image pixels, right: $128 \times 80$ image pixels)

Approved for public release; distribution unlimited.
Fig. 7  Same images as shown in Figs. 5 and 6 but with an even higher resolution (left: 525 x 336 image pixel, right: 3888 x 2492 image pixels). Note that the hoist, surveyor, and engineers wearing hard hats in the far-field of the imaged scene all required increased image resolution to be clearly identified.

Table 1  Image resolution information (in numbers of pixels)

<table>
<thead>
<tr>
<th>Object</th>
<th>Fig. 7 (left)</th>
<th>Fig. 7 (right)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main image</td>
<td>525 x 336</td>
<td>3888 x 2492</td>
</tr>
<tr>
<td>Reactor dome</td>
<td>191 x 51</td>
<td>1028 x 256</td>
</tr>
<tr>
<td>Hard hat</td>
<td>82 x 45</td>
<td>405 x 225</td>
</tr>
<tr>
<td>Danger sign</td>
<td>32 x 36</td>
<td>64 x 69</td>
</tr>
<tr>
<td>Hoist</td>
<td>5 x 24</td>
<td>39 x 175</td>
</tr>
<tr>
<td>Surveyor</td>
<td>5 x 11</td>
<td>34 x 74</td>
</tr>
<tr>
<td>Engineers</td>
<td>5 x 5</td>
<td>32 x 41</td>
</tr>
</tbody>
</table>

Thus, we have visually demonstrated that image resolution and image context play an important role in being able to clearly recognize and identify individual objects for scene understanding. Here, object recognition may be achieved through applications associated with deep learning neural networks.\textsuperscript{56–58} Also, it is important to note that image context information can include elements related to time, geographical location, and environmental conditions,\textsuperscript{12–14} which can help to provide a more detailed description of recorded scenes for indexing and future retrieval, as is discussed next.

4. **Time- and Space-Varying Elements of Scene Understanding**

While image resolution is a key element for identifying objects and quickly discerning the gist and general layout of a recorded scene, there are many other time- and space-varying elements that are equally important for scene understanding that should be addressed from the very beginning of the data collection process.
4.1 Environmental Image Information

There are many key pieces of information that can be identified as new image data are being recorded that are important and accessible, but are usually overlooked or left undocumented. For example, one can readily identify a timestamp (relative to the sun’s angle or relative to a world clock), the global positioning system (GPS) position, the prevailing environmental and weather conditions, and the field of view, depth of view, and image resolution (Table 2). The first group, shown in Table 2, focuses on environmental information, such as the GPS position and altitude above ground level (AGL), prevailing weather, cloud cover, ground and road conditions, and visibility (e.g., fog, smoke, haze, obscurants, or optical turbulence).

Identifying the key environmental and terrain conditions can provide location and geographical context information to help categorize image scenes recorded in diverse regions (e.g., coastal, mountain-valley, desert, forest, urban, rural, ocean, and arctic). Detailed terrain characteristics (e.g., muddy, sandy, gravelly, wet, dry, or icy) and reports of the most current weather conditions available (e.g., rain, snow, fog, or haze) can be retrieved and annotated to help describe images that may be used to support the planning and/or execution of military operations, as mentioned above. Typically, changing weather conditions, cloud cover, and visibility will bring about changes in the illumination of a scene, which can affect image contrast and resolution. Time of day and sun angle information also can be retrieved, which can be useful to indicate when glare, shadows, or silhouettes may cause difficulties for computer vision processes. Also, taking note of optical turbulence conditions is important because these effects can significantly degrade and blur image quality due to spatial smearing.

The second group in Table 2 lists elements related to the camera specifications and the image data measurements themselves (e.g., the spatial and temporal image resolutions, field of view, depth of view, and scene color or shading variations). Together with the environmental information, these elements can be used as a basic building block for detailed scene description and image indexing.
Table 2  Time- and space-varying elements of scene understanding

<table>
<thead>
<tr>
<th>Environmental information</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPS position and altitude AGL</td>
</tr>
<tr>
<td>Location: geographical context</td>
</tr>
<tr>
<td>Timestamp</td>
</tr>
<tr>
<td>Weather conditions, sky and cloud cover</td>
</tr>
<tr>
<td>Sun/moon angle</td>
</tr>
<tr>
<td>Ground/road conditions</td>
</tr>
<tr>
<td>Visibility</td>
</tr>
<tr>
<td>Vegetation</td>
</tr>
<tr>
<td>Buildings, parking lots, people, or crowds</td>
</tr>
<tr>
<td>Image/camera information</td>
</tr>
<tr>
<td>Image resolution</td>
</tr>
<tr>
<td>Pixel size and pixel separation</td>
</tr>
<tr>
<td>Scene color or shading variations</td>
</tr>
<tr>
<td>Field of view and depth of view</td>
</tr>
<tr>
<td>Shutter exposure time</td>
</tr>
<tr>
<td>Time interval between image frames</td>
</tr>
<tr>
<td>Time over which images are captured in a sequence</td>
</tr>
</tbody>
</table>

4.2  Scene Description Indexing

Based on the time- and space-varying elements for scene understanding described above, Table 3 provides a top-level view for scene description indexing (i.e., these are the questions that one should endeavor to address as image data are being recorded so that the data can be best indexed and retrieved for later use). In most cases, the image information (right column) can be annotated based on the camera type, lensing, pixel array, and timing specifications. Also, for example, co-located range finder instrumentation could provide effective depth and field of view measurements for this purpose. Additionally, when communications are available, the most current environmental information available (left column) can be extracted from several accessible resources, such as those shown in Table 4. Obtaining the most current data available is advantageous since environmental conditions (e.g., weather and terrain) can change over very short temporal and spatial intervals. For example, access to data from the Department of Defense (DOD) GPS can provide latitude and longitude or Universal Transverse Mercator (UTM) location and timestamp information, commonly reported as Greenwich Mean Time (GMT) or Coordinated Universal Time (UTC). Also, data from the US Naval Observatory (USNO) can provide precise timing information as well as solar and lunar elevation/azimuth angles. Similarly, terrain and geographical location and context information can be provided by satellite and aerial imagery from the US Army Corps of Engineers, Army Geospatial Center (USACE AGC) for military operations or from public Internet resources such as Google, MapQuest, Bing, and Yahoo Maps.
Table 3  Scene description indexing (top-level view)

<table>
<thead>
<tr>
<th>Environmental information</th>
<th>Image/camera information</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  What is the GPS position of the depicted scene?</td>
<td>2  What is the camera field of view?</td>
</tr>
<tr>
<td>2  What is the altitude (AGL) of the depicted scene?</td>
<td>3  What is the scene depth of view?</td>
</tr>
<tr>
<td>3  What is the timestamp of the recorded image?</td>
<td>4  What is the image spatial resolution (in pixels)?</td>
</tr>
<tr>
<td>4  What are the current weather conditions?</td>
<td>5  What is the camera pixel size?</td>
</tr>
<tr>
<td>5  What is the percent sky/cloud cover?</td>
<td>6  What is the scene color (R, B, G or grayscale)?</td>
</tr>
<tr>
<td>6  What is the sun elevation/azimuth angle?</td>
<td>7  What are the spatial shading variations of the scene?</td>
</tr>
<tr>
<td>7  What is the visibility?</td>
<td>8  What is the camera integration time?</td>
</tr>
<tr>
<td>8  What are the current ground/road conditions?</td>
<td>9  What is the camera shutter variations of the scene?</td>
</tr>
<tr>
<td>9  What vegetation is in the field of view?</td>
<td>10 What is the time interval between frames?</td>
</tr>
<tr>
<td>10 What buildings, parking lots, people or crowds are in the field of view?</td>
<td>What is the starting, ending and total time for the recorded image sequence?</td>
</tr>
</tbody>
</table>

Table 4  Available/accessible environmental image information

1  DOD GPS: Latitude/longitude or UTM, altitude (AGL), GMT, or UTC
2  USNO: Precise time, sun/moon elevation/azimuth angle
3  Terrain and location: USACE AGC — Satellite/aerial imagery and terrain analysis
4  Terrain and location: Google, MapQuest, Bing, Yahoo Maps
5  Weather: USAF 557th Weather Wing
6  Weather: National Weather Service (NWS) and National Centers for Environmental Information (NCEI)
7  Weather: Intellicast, AccuWeather, Weather Underground

Weather conditions and related oceanic, atmospheric, and geophysical data are also available for the military through the US Air Force 557th Weather Wing (i.e., formerly the US Air Force Weather Agency) and for the civilian community through the NWS and NCEI. Daily NWS weather reports can be found online containing hourly records citing the date, time, wind speed (miles per hour), visibility (miles), weather (i.e., rain, snow, fog, haze, etc.), sky/cloud condition (reported as overcast [OVC], broken [BRK], scattered [SCT], or clear [CLR] along with the cloud ceiling height in hundreds of feet AGL), air temperature, dew point temperature, relative humidity (%), pressure, and precipitation (in inches). Naturally, current weather and weather forecast information are readily found on Internet web sites, such as Intellicast, AccuWeather, and Weather Underground. Note however, that in areas where communications are either restricted or unavailable, the information needed to describe the scene, i.e., as outlined in Table 3, should instead be gleaned from the recorded images when they are retrieved for analysis.

Thus, we have shown that abundant time- and space-varying environmental image information can be accessed and annotated to augment image data as they are being recorded for a better organized, top-down approach to scene description, indexing and image retrieval.
5. Summary and Conclusions

In this report, we proposed that it is important to incorporate space- and time-varying environmental image information at the start of the data collection process in order to provide the end user (Soldier) with a better organized, top-down approach to index and retrieve image data for operational use and analysis. We provided several examples to show that space- and time-varying elements of environmental image information (and changes in image context) can be used as a basic building block for detailed scene description, and as such, could be used to support Army mission planning and execution. In conclusion, we anticipate that incorporating space- and time-varying environmental image information in the data measurement process will lead to 1) improved autonomous intelligent systems supporting Army missions in complex and changing environments and 2) improved course of action strategies based on scene understanding (algorithms and analysis) incorporating battlefield environments changing in space and time.
6. References


34. Teutsch M. Moving object detection and segmentation for remote aerial video surveillance [doctoral dissertation]. Karlsruhe (Germany): Karlsruhe Institute of Technology; 2014.


### List of Symbols, Abbreviations, and Acronyms

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGC</td>
<td>Army Geospatial Center</td>
</tr>
<tr>
<td>AGL</td>
<td>above ground level</td>
</tr>
<tr>
<td>BRK</td>
<td>broken</td>
</tr>
<tr>
<td>CLR</td>
<td>clear</td>
</tr>
<tr>
<td>DOD</td>
<td>Department of Defense</td>
</tr>
<tr>
<td>GMT</td>
<td>Greenwich Mean Time</td>
</tr>
<tr>
<td>GPS</td>
<td>global positioning system</td>
</tr>
<tr>
<td>NCEI</td>
<td>National Centers for Environmental Information</td>
</tr>
<tr>
<td>NWS</td>
<td>National Weather Service</td>
</tr>
<tr>
<td>OVC</td>
<td>overcast</td>
</tr>
<tr>
<td>SCT</td>
<td>scattered</td>
</tr>
<tr>
<td>USACE</td>
<td>US Army Corps of Engineers</td>
</tr>
<tr>
<td>USNO</td>
<td>US Naval Observatory</td>
</tr>
<tr>
<td>UTC</td>
<td>Coordinated Universal Time</td>
</tr>
<tr>
<td>UTM</td>
<td>Universal Transverse Mercator</td>
</tr>
</tbody>
</table>