ARL-TR-7419 ● SEP 2015

US Army Research Laboratory

Space and Time Scale Characterization of Image Data in Varying Environmental Conditions for Better Scene Understanding

by Arnold D Tunick

ARL-TR-7419 ● SEP 2015

US Army Research Laboratory

Space and Time Scale Characterization of Image Data in Varying Environmental Conditions for Better Scene Understanding

by Arnold D Tunick
Computational and Information Sciences Directorate, ARL

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* Sep 2015	**2. REPORT TYPE** Final	**3. DATES COVERED (From - To)** 10/2014–07/2015

4. TITLE AND SUBTITLE
Space and Time Scale Characterization of Image Data in Varying Environmental Conditions for Better Scene Understanding

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Arnold D Tunick

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
US Army Research Laboratory
ATTN: RDRL-CII-A
2800 Powder Mill Road
Adelphi, MD 20783-1138

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-7419

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Interpreting spatially and temporally changing scenes due varying environmental conditions and visual motion of objects within the field of view can pose serious challenges for rapid and robust scene understanding, particularly for problems of interest to the Army. Weather events, smoke, obscurants, or other changes in lighting and visibility can all affect image contrast and resolution, as can blurring caused by rapid movements or long exposure times in single frames and sequences of recorded images. To help mitigate some of the difficulties inherent in measuring and analyzing changing scenes, I propose that it is important to focus on the space and time scales of image data from the very beginning of the data collection process. This top-down approach not only helps to systematically characterize the measured data, but can help the end user determine which analysis or computer vision tasks are feasible with the available data. Alternately, this approach may be useful to predetermine what image resolutions are needed to enable more intelligent data collection. In this report, I begin to explore the space and time scale aspects of image data, discuss image motion characterization, and propose follow-on research studies to develop numerical algorithms and experiments to explore and analyze changing image scenes with new or existing data sets.

15. SUBJECT TERMS
computer vision, image resolution, gist, visibility, illumination, visual motion, optical flow velocity, acceleration

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Arnold D Tunick
Unclassified	Unclassified	Unclassified	UU	22	**19b. TELEPHONE NUMBER (Include area code)** (301) 394-2111

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

Acknowledgments

I thank R Meyers, G Warnell, and P David for helpful discussions. This research was supported by the US Army Research Laboratory (ARL).

INTENTIONALLY LEFT BLANK.

1. Introduction

Rapid and robust scene understanding is a critically important goal for enhanced robot autonomy;[1] however, the interpretation of spatially and temporally changing image scenes due to varying environmental conditions can pose serious challenges for computer vision processes, such as those associated with vision-based place recognition and navigation.[2,3] Such challenges also can extend to interpreting changing scenes due to visual motion of objects within the field of view.[4] As Tu et al.[4] discuss, adverse weather or illumination conditions can make the appearance of moving objects unclear, so that identifying moving objects in outdoor environments becomes more difficult for robot vision systems. As an example, time- and space-dependent environmental effects on image contrast and resolution can be brought about by rain and snow weather events, fog, smoke, obscurants or other changes in lighting and visibility.[5] Alternately, visually degraded or blurred images due to the motion of objects can occur due to rapid movements or long exposure times in both single frames and sequences of recorded images.[2,6]

With regard to lighting variations, Andreev et al.[7] described a method to estimate the effect of space and time changes in scene illumination on the optical flow field in a movie. Their research[7] is unique, because there have been many optical flow approaches used to detect motion of objects in a scene that do not have the scope of space and time scaling and analysis, even though they may be helpful in a variety of applications.[8–10] Also, camera motion may introduce some unmanageable artifacts with the gradient-based optical flow approach if it is not augmented by more sophisticated spatio-temporal analyses.[11–13] Other difficulties in image motion analysis can arise if objects in the scene have reflections; when new objects appear or old ones disappear; or when describing transparent motions, for example, the motion of objects behind smoke, foliage, or a fence.[13] Here, in addition to capturing the motion of individual objects, it appears necessary to capture the relative motion between individual objects and the time and space resolutions of the information being collected.

To help mitigate some of the difficulties associated with the measurement and analysis of changing scenes, I propose that it is important to consider the space and time scales of image data from the very beginning of the data collection process. Incorporating key space and time scale information at the time of recording not only helps to systematically characterize the measured data but can provide the future analyst with a top-down approach to determine what analysis or computer vision tasks are feasible with the available data. This kind of enhanced analysis and decision making may also be applied to future autonomous systems. Alternately, if an image analysis or computer vision objective is pre-known then it would be useful to determine what image resolutions are needed to enable more intelligent data collection. If neglected, the end user advantages provided by space and time scale characterization may be inextricably lost.

In this report, I begin to explore the space and time scales of image data as they are related to the measurement and analysis of changing image scenes, and whether scene variations are due to environmental conditions or the motion of objects within the field of view or both.

2. Space and Time Scales

2.1 Primary Space and Time Scales

This section provides a framework to help categorize the spatial and temporal properties of image data. Relevant time scales include, but are not limited to, the shutter exposure time, the time interval between frames, time over which images are captured in a sequence, and the 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, color matrix size, 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, which can be used to describe the various spatial and temporal resolutions of objects and/or activities in a recorded image scene. Here, Δs and Δt represent changes in position and time, respectively.

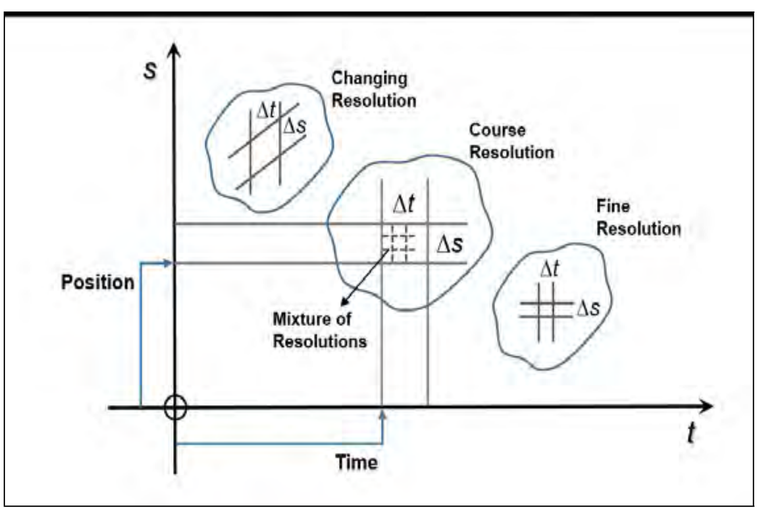

Fig. 1 Primary space (s) and time (t) scales

2.2 Image Resolution and Field of View

To begin to demonstrate the impact of varying image resolution and field of view on scene analysis, try to identify the 3 dome shapes shown in Fig. 2. Without some additional information related to the object size, texture, or shape in relationship to other objects that may be visible 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 numbers of pixels that comprise the image data input. Interestingly, Torralba[14] reported that for human vision the brain can comprehend the gist of an image scene remarkably quickly, whether low resolution 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.

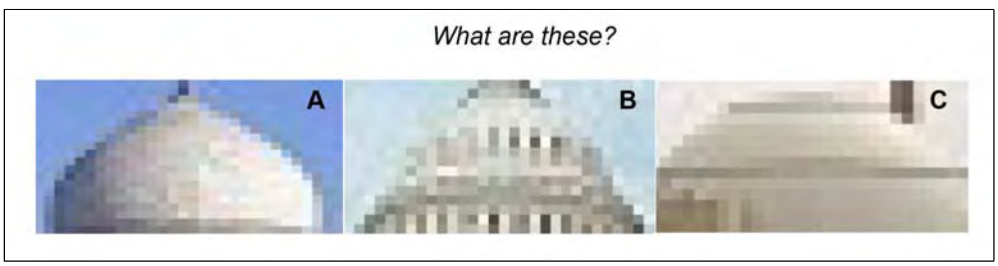

Fig. 2 Can you correctly identify these images? Image resolution: a) 30 x 20 pixels, b) 30 x 14 pixels, and c) 30 x 16 pixels.

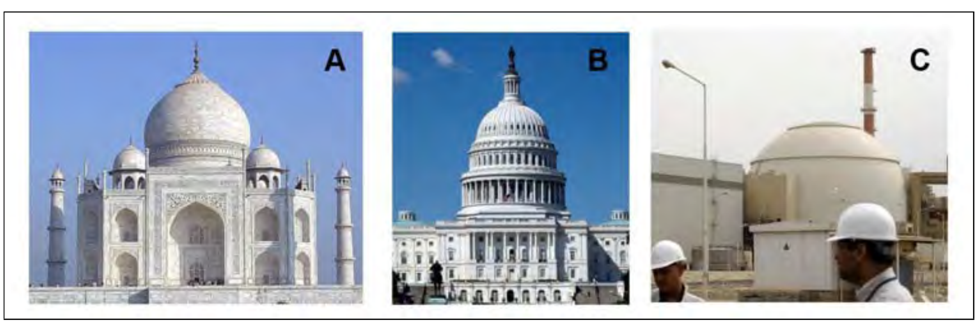

Fig. 3 Higher resolution image scenes corresponding to the 3 shapes shown in Fig. 2. a) photo: Taj Mahal (desktopdress.com), b) photo: US Capitol Dome (Library of Congress), and c) photo: nuclear power plant, Bushehr, Iran (Behrouz Mehri/AFP/Getty Images).

Let's examine image resolution more closely. Can you identify the 2 extracted 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.

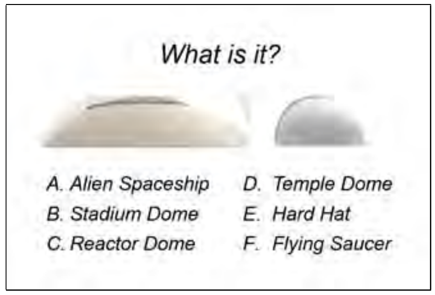

Fig. 4 Can you correctly identify these objects?

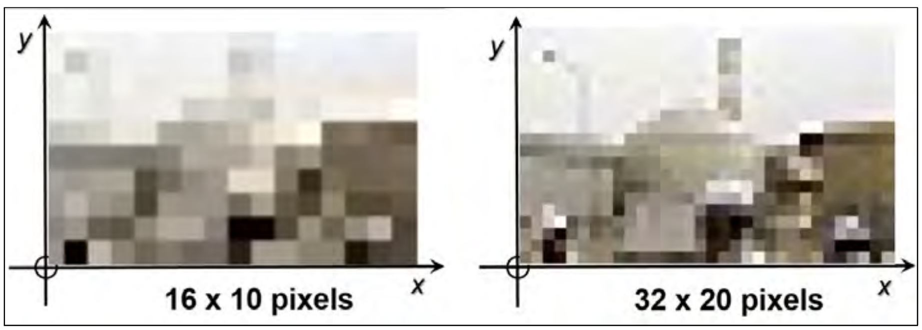

Fig. 5 Low resolution images of the scene from which the objects in Fig. 4 were taken, where neither large nor small objects are discernible. (Left: 16 x 10 image pixels, Right: 32 x 20 image pixels.)

Imagine you are in new surroundings and you only have a low resolution, e.g., 32 x 32 pixels, imaging camera that can be used for wide area coverage. However, you can change lenses on the camera to narrow your field of view such that a region of the scene that was previously imaged by a single pixel is now imaged with 32 x 32 pixels (Fig. 6). If applied to every pixel location in the original image matrix, the image resolution of the wide area view would be increased to 1024 x 1024 pixels. However, this would require much more time and data collection on your part, and large amounts of data transmission is often costly, i.e., bandwidth limited. Given that you are likely to be more selective in recording the narrower fields of view, e.g., focusing on an identified area of interest, it would be helpful to develop some useful strategies to enable rapid and robust scene analysis. As an example, Warnell et al.[15] recently discussed concepts associated with visual saliency to enable enhanced camera control for tasks such as automatic navigation and scene exploration. Saliency estimation, which is a computational identification of various elements in a scene that are likely to catch the attention of a human observer, is a valuable tool in image analysis and processing.[16] Nevertheless, future work may include extending these concepts to interpreting changing (dynamic) scenes,[15] such as those described above.

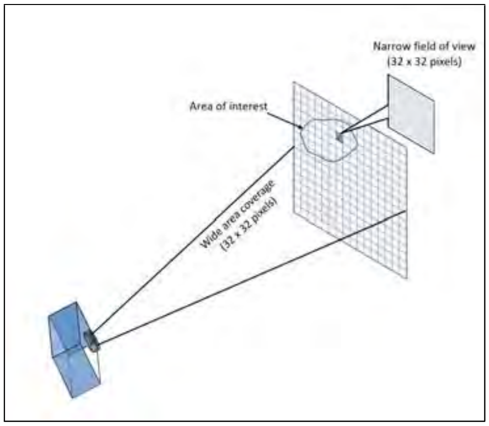

Fig. 6 Framework for follow on numerical experiments to focus on scene analysis strategies

Such a problem can be investigated and characterized through a series of human-in-the-loop numerical experiments. For example, it may be useful at first to develop a general object searching algorithm that focuses on the narrow fields of view at randomly selected locations within an identified area of interest. Then, one can begin to analyze how many randomly revealed narrow fields of view would be needed to clearly identify specific object shapes and textures or to capture the gist of the image scene (as discussed by Torralba).[14] Later on, this approach can be expanded using a different or improved strategy.

Of course, the degree of image resolution needed for a particular task depends on the analysis or computer vision problem of interest.[2,17–22] For example, with regard to image analysis and labeling, compare the low resolution images in Fig. 5 to the slightly higher resolution images shown in Fig. 7. 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 analyzed (Fig. 8) 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. 8 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., ≥32 x 32 pixels, to be clearly identified (visually compare right vs. left in Fig. 8). Table 1 provides the image resolution details in numbers of pixels for these labeled objects.

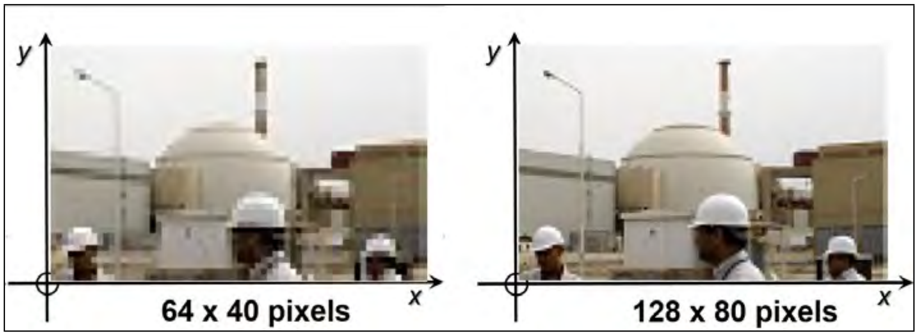

Fig. 7 **Same images as shown in Fig. 5 but with slightly higher resolution. Left: 64 x 40 image pixels. Right: 128 x 80 image pixels.**

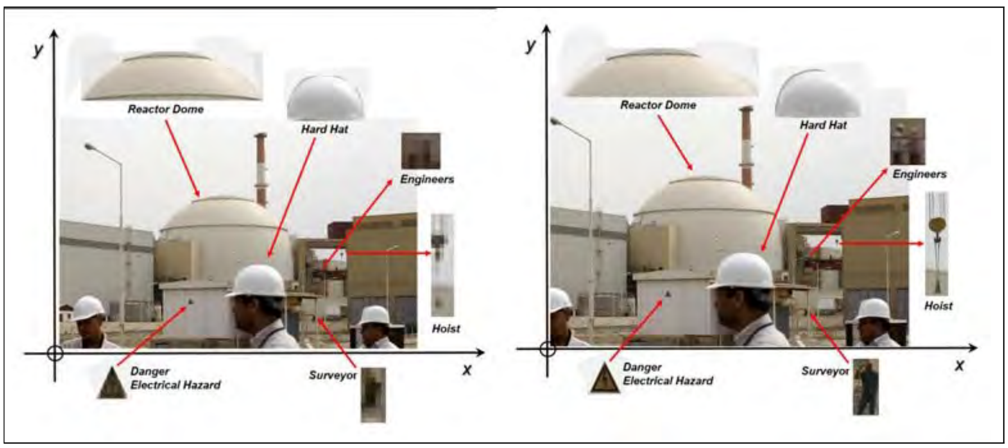

Fig. 8 **Same images as shown in Figs. 5 and 7 but with even higher resolution. Left: 525 x 336 image pixels. Right: 3888 x 2492 image pixels. Note that the hoist, surveyor and engineers wearing hard hats in the far-field of the imaged scene all required increased image resolution to be clearly identified.**

Table 1 **Image resolution information (in numbers of pixels)**

	Fig. 8 (Left)	**Fig. 8 (Right)**
Main image	525 x 336	3888 x 2492
Reactor Dome	191 x 51	1028 x 256
Hard Hat	82 x 45	405 x 225
Danger Sign	32 x 36	64 x 69
Hoist	5 x 24	39 x 175
Surveyor	5 x 11	34 x 74
Engineers	5 x 5	32 x 41

In this section, we have shown an example of varying image resolution as it relates to detailed analysis and labeling of this outdoor scene. In follow-on research, we can similarly explore the impact of varying time scale related image resolution on scene analysis using sequences of recorded images.

3. Image Data Input

There are several key pieces of information that can be identified as new image data are being recorded that are important and accessible, but usually overlooked or left undocumented. For example, one can readily identify a timestamp, the global positioning system (GPS) position, the prevailing environmental and weather conditions, the field of view, depth of view, and image resolution, as noted above. Table 2 provides a list of several space and time scale dependent elements that can be incorporated with measured image data. The first group focuses on environmental effects, such as the weather, cloud cover, ground and road conditions, and visibility. Identifying environmental conditions is an effective way to categorize diverse data sets of image scenes for later use and analysis. For example if, at a later time, an end user needs to find image data with a certain resolution in raining or low-light conditions, then incorporating the space-time related elements listed in Table 2 at the time of recording the data can provide the desired benefit.

In addition, changing environmental conditions can affect image contrast and resolution due to weather events and changes in visibility or cloud cover and these effects often coincide with lighting changes in an imaged scene, e.g., those due to increased scattering and attenuation of light in adverse weather conditions.[5] Time of day and sun angle information can be useful also to highlight conditions when increased glare, shadows, or silhouettes can cause difficulties for image analysis and computer vision related processes.[20,23,24] Taking note of fog, smoke, obscurants, and optical turbulence conditions is also important because these effects can significantly degrade image quality due to spatial smearing of shapes, textures, and moving elements in an imaged scene[25] Similarly, identifying ground and road conditions, e.g., wet or dry, icy, sand, or gravel, can be used in subsequent image analyses or can support autonomous systems with regard to navigation tasks, tracking personnel, or detecting vehicles.

The second group in Table 2 lists elements related to the image data measurements themselves, e.g., the spatial and temporal image resolutions, field of view, and depth of view. Together with the environmental effects, these data can be used as a basic building block for the analysis of changing image scenes.

Table 2 Incorporating key space and time scale related information as image data are being collected

Environmental Effects
Weather conditions (rain, snow, haze, fog, or hail)
Sun angle, sky, and cloud cover
Ground/road conditions (dry, wet, icy, sand, gravel, rocky, etc.)
Visibility (fog, smoke, obscurants, or optical turbulence)
Image Data Measurements
GPS position
Timestamp (relative to sun angle or relative to a world clock)
Image resolution
Pixel size and pixel separation
Field of view and depth of view
Shutter exposure time and time interval between image frames
Time over which images are captured in a sequence

4. Image Motion Characterization

Polana and Nelson[26] suggested that image motion in a scene can be categorized into 3 parts. The first group of motions are those having statistical regularities, i.e., they are repeatable in both space and time, such as the action of water waves or the motion of clouds, trees, and leaves. In contrast, the second group consists of activities, repeatable over time but not over space, such as people walking, biking, or talking. The third group includes motion events[11] that are not repeatable in either space or time, such as a person throwing a ball or entering a room. According to Laptev,[11] such events correspond to features in the image scene appearing or disappearing and with non-constant motion, which often correspond to changes or discontinuities in velocity and acceleration. An alternate method for motion event analysis based on visual attention and temporal salience has been discussed by Thomas.[27]

In follow-on research, we can develop additional numerical algorithms and experiments that can be implemented using new or existing data sets to help recognize the motion of individual objects in a scene and mitigate any undesirable artifacts due to camera motion.[12] Briefly, we can express the optical flow (i.e., image gradient) velocity ($\vec{u}$) of a moving object in a scene as

$$\vec{u} = \frac{dx_i}{dt} = \vec{v}\,(\vec{x_i}, t_m) + \vec{w}, \tag{1}$$

where, $\vec{v}$ is the velocity of an individual element in the image, which is a function of position, $\overrightarrow{x_1} = x_1 + x_2 + x_3$, and time (of successive frames), $t_m = t_1, t_2, t_3$... Also in Eq. 1, $\vec{w}$ is the velocity of the camera motion from one image frame to the next, which in some cases is considered a constant. Next, we define the divergence of the optical flow velocity as

$$\nabla \vec{u} = \frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3}, \tag{2}$$

and realize from Eqs. 1 and 2 that if one adds a constant ($\vec{w}$) to the velocity field, then the divergence of the velocity field ($\nabla \vec{u}$) remains unchanged. We can also explore the optical flow acceleration ($\vec{a}$) and its divergence ($\nabla \vec{a}$) in a similar manner. We anticipate that our research results will provide many useful insights toward developing novel strategies for the analysis of space and time varying scenes.

5. Summary and Conclusions

In this report, I began to explore the space and time scale aspects of image data as they are related to the measurement and analysis of changing image scenes, and whether scene variations are due to environmental conditions or the motion of objects within the field of view or both. I showed an example that demonstrated the impact of varying image resolution on the detailed analysis and labeling of an outdoor scene. I also provided a list of several space and time scale dependent elements that, if incorporated at the start of the image data measurement process, can provide an end user with a better organized, top-down approach to determine what analysis or computer vision tasks are feasible with the available data. Finally, I discussed image motion characterization and proposed a follow-on research study to develop numerical algorithms and experiments to explore and analyze changing image scenes with new or existing data sets. I anticipate that this research will help to advance Army relevant technologies in scene understanding for enhanced robot autonomy.

6. References

1. System intelligence and intelligent systems, US Army Research Laboratory S&T Campaign Plans 2015–2035, 98–99, 2014.

2. Milford M. Vision-based place recognition: How low can you go? Intl J Robotics Research. 2013;32(7):766–789.

3. Milford M, Vig E, Scheirer W, Cox D. Vision-based simultaneous localization and mapping in changing outdoor environments. J Field Robotics. 2014;31(5):780–802.

4. Tu Z, Zheng A, Yang E, Luo B, Hussain A. A biologically inspired vision-based approach for detecting multiple moving objects in complex outdoor scenes. Cognitive Computation, DOI 10.1007/s12559-015-9318-z, 2015.

5. Narasimhan SG, Nayar SK. Vision and the atmosphere. Intl J Computer Vision. 2002;48(3):233–254.

6. Kim MD, Ueda J. Dynamics-based motion de-blurring for a PZT-driven, compliant camera orientation mechanism. Intl J Robotics Research. 2015;34:653–673.

7. Andreev R, Scherzera O, Zulehner W. Simultaneous optical flow and source estimation: Space—time discretization and preconditioning. Applied Numerical Mathematics. 2015;96:72–81.

8. Fortun D, Bouthemy P, Kervrann C. Optical flow modeling and computation: A survey. Computer Vision and Image Understanding. 2015;134:1–21.

9. Fleet DJ, Weiss Y. Optical flow estimation. Chapter 15 in Mathematical Models in Computer Vision: The Handbook, 239–258, Springer, 2005.

10. Barron J, Fleet D, Beauchemin S. Performance of optical flow techniques. Intl J Computer Vision. 1994;12(1):43–77.

11. Laptev I. Local spatio-temporal image features for motion interpretation. Doctoral Thesis, ISBN: 91-7283-793-4, KTH Royal Institute of Technology, Stockholm, Sweden, 2004.

12. Fagerstrom D. Galilean differential geometry of moving images. chapter in Computer Vision - ECCV 2004, 494–506, Springer-Verlag, Berlin, Heidelberg, 2004.

13. Fagerstrom D. Spatio-temporal scale-space theory. Doctoral Thesis, ISBN: 978-91-7501-024-3, KTH Royal Institute of Technology, Stockholm, Sweden, 2011.

14. Torralba A. How many pixels make an image? Visual Neuroscience. 2009;26:23–131.

15. Warnell G, David P, Chellappa R. Ray saliency: Bottom-up visual saliency for a rotating and zooming camera. Intl J Computer Vision, DOI 10.1007/s11263-015-0842-9, 2015.

16. Perazzi F, Krahenbuhl P, Pritch Y, Hornung A. Hornung, Saliency filters: Contrast based filtering for salient region detection. IEEE Conference on Computer Vision and Pattern Recognition; June 16–21, 2012, Providence, RI.

17. Fleuret F, Geman D. Coarse-to-fine face detection. Intl J Computer Vision. 2001;41:85–107.

18. Park D, Ramanan D, Fowlkes C. Multiresolution models for object detection. chapter in Computer Vision - ECCV 2010, 241–254, Springer-Verlag, Berlin, Heidelberg, 2010.

19. Pedersoli M, Vedaldi A, Gonzalez J. A coarse-to-fine approach for fast deformable object detection. IEEE Computer Vision and Pattern Recognition (CVPR) Conference; 2011 June 21–23; Colorado Springs.

20. Olszewska JI. Semantic, automatic image annotation based on multi-layered active contours and decision trees. Intl J Advanced Computer Science and Applications. 2013;4(8):201–208.

21. Karpathy A Fei-Fei L. Deep visual-semantic alignments for generating image descriptions. arXiv:1412.2306v2, 2015.

22. Wigness M, Draper BA, Beveridge JR. Efficient label collection for unlabeled image datasets. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015 4594–4602; Boston, MA.

23. Shafer SA. Shadows and silhouettes in computer vision. Kluwer Academic Publishers, 1985.

24. Reddy D, Veeraraghavan A. Lens flare and lens glare. in Computer Vision: A Reference Guide, 445—447, Springer, 2014.

25. Roggemann MC, Welsh BM, Hunt BR. Imaging through turbulence. CRC press, Boca Raton,1996.

26. Polana R, Nelson R. Temporal texture and activity recognition, chapter in Motion-Based Recognition, Kluwer Academic Publishers, Dordrecht, Boston, London, 87–124, 1997.

27. Thomas G. Visual attention framework: Application to event analysis. Doctoral Thesis, Vrije Universiteit Brussel, Brussels, Belgium, 2009.

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
2 (PDF)	DIRECTOR US ARMY RESEARCH LAB RDRL CIO LL IMAL HRA MAIL & RECORDS MGMT
1 (PDF)	GOVT PRINTG OFC A MALHOTRA
7 (PDF)	DIRECTOR. US ARMY RESEARCH LAB RDRL CII B BROOME RDRL CII A S YOUNG A TUNICK G WARNELL P DAVID RDRL CIN T R MEYERS K DEACON

INTENTIONALLY LEFT BLANK.