available at www.sciencedirect.com

SCIENCE DIRECT®

journal homepage: www.elsevier.com/locate/diin

Digital Investigation

Detecting false captioning using common-sense reasoning

*Sangwon Lee, David A. Shamma, Bruce Gooch**

Department of Electrical Engineering and Computer Science, Northwestern University, 2145 Sheridan Road, Technological Institute L359, Evanston, IL 60208, USA

Keywords:
Image forgery detection
False captioning
Digital photomontage
Image segmentation
Common-sense reasoning

ABSTRACT

Detecting manipulated images has become an important problem in many domains (including medical imaging, forensics, journalism and scientific publication) largely due to the recent success of image synthesis techniques and the accessibility of image editing software. Many previous signal-processing techniques are concerned about finding forgery through simple transformation (e.g. resizing, rotating, or scaling), yet little attention is given to examining the semantic content of an image, which is the main issue in recent image forgeries. Here, we present a complete workflow for finding the anomalies within images by combining the methods known in computer graphics and artificial intelligence. We first find perceptually meaningful regions using an image segmentation technique and classify these regions based on image statistics. We then use AI common-sense reasoning techniques to find ambiguities and anomalies within an image as well as perform reasoning across a corpus of images to identify a semantically based candidate list of potential fraudulent images. Our method introduces a novel framework for forensic reasoning, which allows detection of image tampering, even with nearly flawless mathematical techniques.

1. Introduction

While 2D images have been a powerful media for delivering and communicating information, some creators of the images have been tempted to tamper the original in order to distort the reality for their particular interests. This includes yellow journalists who want to make up their own stories, photojournalists who are searching for dramatic scenes, scientists who forge or repeat images in academic papers, and politicians who try to direct people's opinion by exaggerating or falsifying political events. These image forgery examples serve different purposes. Already abundant, they are becoming more clever and even diverse with the advent of digital image editing software.

In *Photo fakery*, Brugioni (1999) argues photo manipulation techniques fall into four categories:

- *Deletion of details*: removing scene elements
- *Insertion of details*: adding scene elements
- *Photomontage*: combining multiple images
- *False captioning*: misrepresenting image content

In this paper, we discuss techniques for detecting manipulations related to *photomontage* and *false captioning* which result in duplicated scene elements from a set of images in a single image (photomontage) or images in which the content has been altered prior to image creation (false captioning). Fig. 1 shows an example of photomontage before the digital age.

* *Corresponding author.*
E-mail addresses: s-lee@cs.northwestern.edu (S. Lee), ayman@cs.northwestern.edu (D.A. Shamma), bgooch@cs.northwestern.edu (B. Gooch).
1742-2876/$ – see front matter © 2006 DFRWS. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.diin.2006.06.006

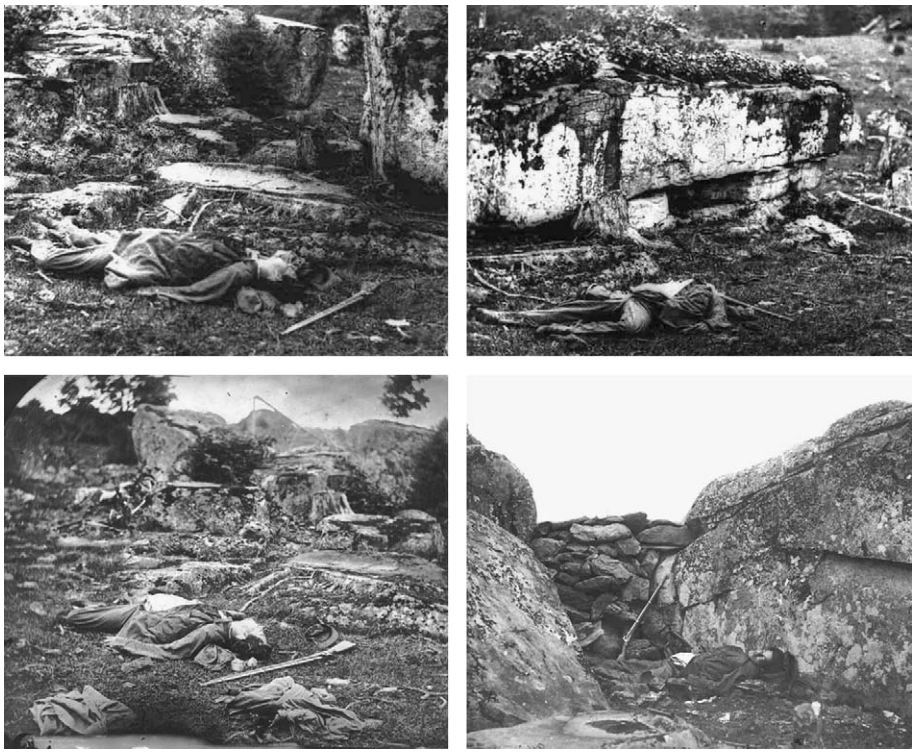

Fig. 1 – A series of photos of the "Devils Den" sniper photographed after the battle of Gettysburg. The first three photos show the soldier where he fell in battle. The fourth show the body "posed" for dramatic effect.

We propose using the artificial intelligence (AI) techniques of common-sense reasoning to detect duplicated and anomalous elements in a set of images. For many images, the key content is a small set of objects, which makes the common-sense reasoning more tractable due to the problem's scale. The premise of our method is that if the key objects in a set of images can be identified, we can use common-sense reasoning to determine if objects have been moved, added, or deleted from a scene prior to image creation. The technique has three basic stages: image segmentation, segment classification, and reasoning. Fig. 2 summarizes the algorithm. In this paper, we propose several techniques which are applicable to each stage.

1.1. *Segmentation*

We first segment the source image into regions. We then use an importance map to select a set of regions of importance (ROI) to compare across images in a given corpus. Alternatively, the algorithm can be applied in a semi-automatic fashion by having a user specify the ROI. In Section 3.1, we discuss the techniques involved in segmenting the image and combining adjacent regions based on their spatial distribution of color/intensity. In order to identify important regions, we create an importance map of the source image using saliency and face detection.

1.2. *Classification*

Image classification is perhaps the most important part of digital image analysis. While a false colored image illustrating image features is nice, it is useless to an investigator unless they know what the colors mean. We propose a segment based classification scheme. The benefits of segment-wise classification are twofold. First, brute-force pixel comparison is unmanageably slow; segment-wise classification reduces the size of the problem-space significantly by highlighting ROI for comparison. Second, comparing the relationship of

Fig. 2 – Results of mean-shift segmentation with parameters $h_s = 7$, $h_r = 6$, and $M = 50$ (top). Results of region merging (bottom).

the ROI across a corpus of images gives us the ability to determine if a scene has been manipulated during the photo recording process and to automatically reason about any post processing of a given image.

1.3. *Common-sense reasoning*

We propose two common-sense reasoning approaches to assist digital forensics. First, to resolve local classification ambiguities within images, we will query a knowledge base to resolve the proper relation. Second, we reason across a larger corpa of images to find unqiue or missing elements during an investigation. The reasoner's focus is on the ROI components from the previous classification stage, which provides a more tractable space for the reasoner to operate.

2. Related work

Brugioni (1999) realizes that image forgery is a long-existed problem that goes back to the birth of photography. He describes various traditional forgery techniques in different fields: political campaign, delivering false military information, falsifying proofs in lawsuit cases, and UFO and ghost pictures. He also includes several methods for spotting those fakery: examining lights and shadows, perspective, depth of field, discontinuous lines, and physically impossible contents of the scene.

As digital imaging becomes a dominant way of recording and preserving photos, previous image synthesis techniques also became rapidly available in the digital domain. Thus to general users, as a side effect, we see significant problems arising in different areas that have not seen the problem before; academia is one of them (Wade, 2006). In spite of their willingness to adopt forgery detection techniques, the problem mainly remains unsolved and depends largely on the author's integrity.

A handful of researchers suggested signal-processing based techniques that can solve some of these problems. Popescu and Farid (2005) detect image resampling, which occurs when the original picture is resized, rotated, or stretched. Expectation–maximization algorithm finds correlation between neighboring image pixels introduced by resampling.

Finding repetition in image segments is a useful technique to identify copy–paste image manipulation (Wade, 2006). The previous approaches exhaustively match the suspected image block against the divided blocks of the entire image. Discrete Cosine Transform (Popescu and Farid, 2004) or Principle Components Analysis (Fridrich et al., 2003) is applied to find the mathematical representations of each image block and is used for the comparison. While this conversion improves the robustness from the naive pixel value matching, they have two significant limitations: the matching should be exhaustive which prohibits comparison matching of large image database, and the block size of the original and copied image segment should be same; it cannot detect repetition of the scaled or rotated image block.

In order to detect the photomontage, Ng et al. (2004) use a higher-order statistics using the fact that a composite signal from different spliced images is likely to introduce discontinuities. Popescu (2005) examined local noise level variance to find the splice boundary. However, state-of-the-art image compositing techniques in computer graphics community like, gradient-based digital photomontage techniques Perez et al., 2003 or graph-cut base algorithms Agarwala et al., 2004 can automatically choose the optimal boundary path, which makes the detection with these methods much harder.

Different lighting direction is a good indicator of different image sources. While Johnson and Farid (2005) estimate the direction of the lighting from a single image, the assumption—the normal direction of the object being examined should be known—is too restricting.

Vast research on digital watermarking provides various ways to protect the authenticity of a digital data (Ferrill and Moyer, 1999). However, the major drawback of digital watermarking is that the watermark should be embedded to the image before any tampering. The problem of protecting images from fake watermarking has not been solved completely.

In a related task, the AI community is facing similar issues. While their concerns are not image forgery per se, much work in AI deals with identifying relationships and stories in scenes and scenarios. Here, common-sense reasoning is used to resolve spatial relations across segments within images (Tomai et al., 2004) and identify information that is missing from intelligence scripts (e.g. kidnapping or terrorist scenarios) (Forbus et al., 2005). Throughout our approach, we introduce new semantics to overcome the existing problems within image forensics.

3. Method

3.1. *Image segmentation*

In order to identify the important objects in an image, we must first segment the image. We use mean-shift image segmentation (Meer and Georgescu, 2001) to decompose an image into homogeneous regions. The segmentation routine takes as input, the parameters: spatial radius h_s, color radius h_r, and the minimum number of pixels M that constitute a region. As with other segmentation methods, choosing parameter values is often difficult. Therefore, we over-segment the image using low values of h_r and M and merge adjacent regions based on color and intensity distributions in CIE-Luv. Fig. 2 illustrates an example of this technique. We create a dual graph to store the segmented image regions. Nodes in the dual graph correspond to regions in the segmented image, while edges indicate adjacency. Each node contains a color histogram of CIE-Luv components. Region merging is accomplished by combining adjacent nodes using a color similarity metric proposed by Swain and Ballard (1991).

3.2. *Importance map*

Truly understanding what is important in an image requires a thorough understanding of what the image contains and what the viewer needs. Some recent results suggest that some heuristics work well (albeit imperfectly) on a broad class of images. The two heuristics that are used in most (if not all) systems determining importance in imagery are: specifically

recognizable objects (faces) are usually important; and regions of the image that are most likely to attract the low-level visual system are likely to be important.

To identify the *ROI*, we first compute an importance map that assigns a scalar value to each pixel estimating the importance of that image location based on an attention model. Like previous methods (Suh et al., 2003; Chen et al., 2003), we use measures of visual salience (e.g. image regions likely to be interesting to the low-level vision system) and high-level detectors for specific objects that are likely to be important, such as faces and signs. Our implementation computes the importance map as a scaled sum of a visual saliency algorithm (Itti et al., 1998) and a face detection algorithm (Niblack et al., 1993).

3.3. *Calculating regions of importance (ROI)*

The saliency and face detection algorithms take color images as input return gray-scale images whose pixel values represent the importance of the corresponding pixel in the input image. The importance map, which is the attention model for the image, is built up by combining a series of importance measures. This allows the system to be adapted to differing image creation goals, and to be easily extensible. The importance map computation can accommodate other attention models as desired. A semi-automatic version of the algorithm can be easily implemented by allowing a user specify important regions. We normalize pixel values from the attention model output images and sum them; then re-normalize to create the combined importance map.

Rather than using an exhaustive search to find the best possible regions of importance (ROI), we propose a greedy algorithm. Using the combined importance map, our method finds initial candidate ROI and grow them until they meet the requirements for being the final ROI of the image. The process is described in two steps as follows:

1. *Identify candidate ROI*: a candidate ROI is defined as a minimal region that identifies key important parts of the image. Each importance value from the importance map is mapped to the segmented regions of the image. In order to do this, we calculate an importance value for each node of the DualGraph by summing pixel values in the corresponding region of the importance map.
2. *Grow the ROI*: we extend the method of Swain and Ballard (1991) to include the additional dimension of importance and grow the ROI by combining nodes in the DualGraph. The candidate ROI grows by using a clustering algorithm that considers the unexplored or unattached node with the highest importance. Regions with small importance that are adjacent to regions with higher importance, but which cannot be combined because of color differences, are treated as unimportant. The clustering algorithm is applied recursively until all nodes have been explored. Examples of importance maps and ROI are shown in Fig. 3.

3.4. *Classification*

With supervised classification, one can identify predetermined objects of interest in an image. These are called ''training sites''. An image processing software system (Fig. 4) is then used to develop a statistical characterization of each object. This stage is often called ''signature analysis'' and may involve developing a characterization as simple as the average pixel value of a region, or as complex as detailed analyses of the mean, variances and covariance over all color channels. Once a statistical characterization has been achieved for each object of interest, a given ROI can then be classified by making a decision about which of the signature regions it resembles most.

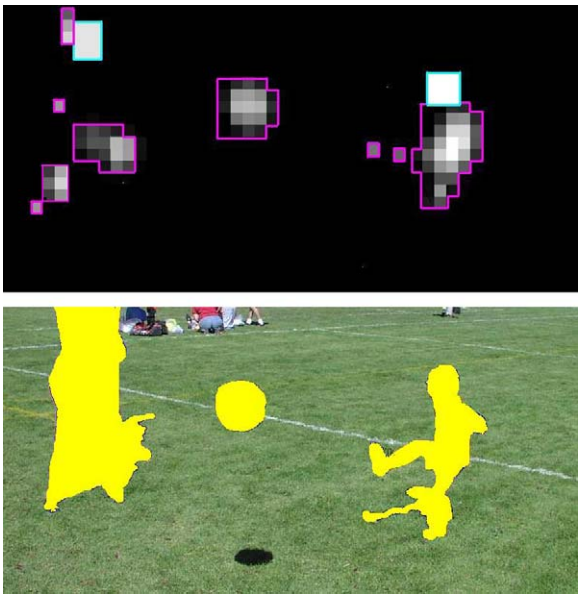

Fig. 3 – Importance map. Saliency regions are outlined in magenta, face regions in cyan (top). Regions of importance (bottom).

Unsupervised classification is a method that examines a large number of ROI and divides them into a number of classes based on groupings present in their statistical characterization. Unlike supervised classification, unsupervised classification does not require analyst-specified training data. The basic premise is that regions within a given type should be close together in the measurement space (i.e. have similar shapes or similar color statistics), whereas regions in different classes should be comparatively well separated. Thus, in the supervised approach an analyst defines useful object categories and the software then examines a given set of images to find similar regions; in the unsupervised approach the computer determines separable region classes, and an analyst then defines their information value.

Unsupervised classification is becoming increasingly popular. The reason is that there are now systems that use clustering procedures that are extremely fast and require little in the nature of operational parameters. Thus it is becoming possible to conduct image analysis with only a general familiarity of the subject matter.

3.5. *Reasoning*

Our first approach with reasoning resolves local ambiguities within an image. In previous work, ambiguities have been

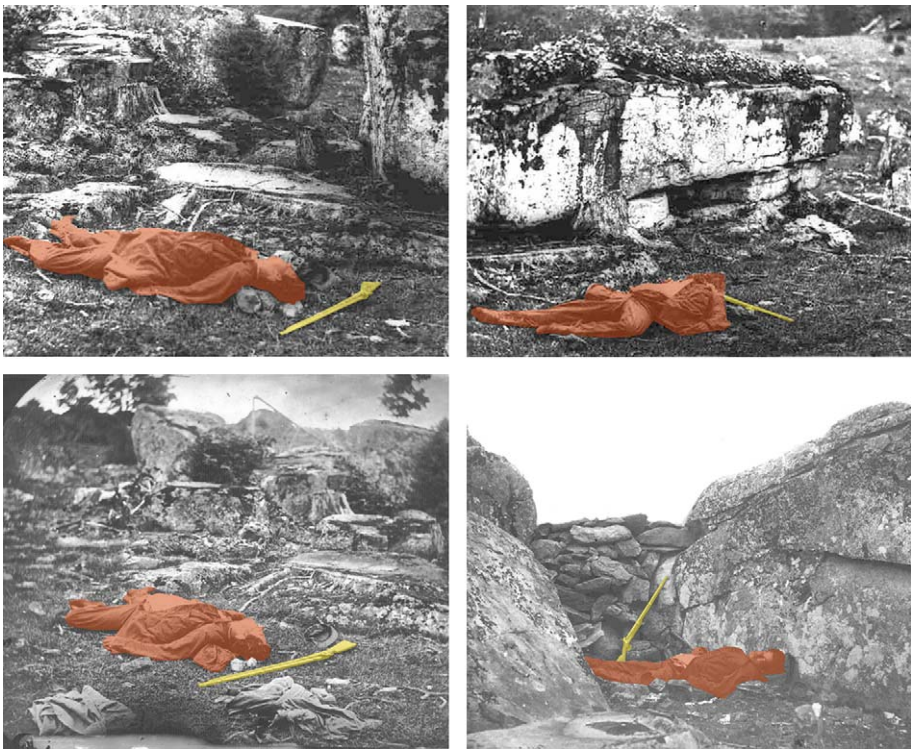

Fig. 4 – We suggest a software system based on a combination of existing tools to identify common objects across a corpus of images. By visualizing such objects, as in this figure, even a layperson can quickly determine whether a given image is falsely captioned.

resolved through hardcoding facts and relations between identified segments (Chen et al., 2005). While this proves effective on a case-by-case basis, we turn to AI techniques for the more general and adaptive solution. A common-sense knowledge base (KB), such as Cyc (Lenat, 1995) and OpenMind (Singh et al., 2002), is well suited for this disambiguation task. For example, given two large, horizontal blue regions in an image, many classifiers cannot distinguish which segment is 'sky' and which is 'water'. A common-sense KB can be queried to find the answer FALSE to the question: 'is the water above the sky?'

Once the ambiguities within the images have been resolved, we then turn to reasoning about the various segments and how they are spatially positioned. Using Cohn's (1996) RCC8 (a calculi for qualitative spatial reasoning), we can reason about the spacial relationship between segments. These qualitative relationships determine if two segments are disjoint, partially tangential, overlapping, contained, etc. With a KB, we can look at an image, find two adjacent regions and determine the likelihood of truth for the image. In an example, we can find information such as 'planes are usually on runways or in the air' or 'buildings are larger than cars.' This reasoning can be done independently across the salient segments (as found in the previous section) as well as across the less salient items—allowing our reasoner to look for foreground or background anomalies. Moreso, as the segmentation and classification methods improve or change, the reasoner can remain in place and simply grow to account for new identifiably segment types.

In many cases, the single image might not tell the complete story. A collection of photos, on the other hand, does show a narrative to a larger story. A collection of photos and facts about the qualitative structure of those photos provides an excellent framework for reasoning across a corpus. For example, a man-made item, such as a plane, in a field of grass should raise suspicion. Unless all the photos in the corpus have similar qualitative structure. Within such a unique corpus, the photo of a plane on a runway should trigger suspicion.

4. Discussion

Our methods are limited by the performance of the components that we use: image segmentation and the importance model must succeed at identifying the important objects. If the performance of these components is insufficient, a semi-automatic version of our method can be applied where the user manually identifies the important object. Our method may be ill suited to images for which this is not the case, for example, when there are too many (or no) distinct, important objects, or when the relationships between the objects are significant.

Our approach introduces a hybrid method for image forensics. Given a subset of a corpus as a suspicious candidate set, we can now analyze the candidates through specific metrics that are optimized to find fakery given the image's qualitative classification. This use of common-sense reasoning goes beyond previous image classification which relied on image metadata (Lieberman and Liu, 2002). In future work, we plan

to integrate the facts discovered in a photo corpus to help identify what evidence may be missing as well as what fact might be unique to this scenario.

Unlike previous signal-processing based approaches, we suggest a framework that can detect anomalies in the contents of the images using common-sense reasoning. While signal-processing based methods are an effective way to finding image additions and deletions. Signal-processing approaches are ill suited for the problem of false captioning.

REFERENCES

Agarwala A, Dontcheva M, Agrawala M, Drucker S, Colburn A, Curless B, et al. Interactive digital photomontage. ACM Transactions on Graphics 2004;23(3):294–302.

Brugioni DA. Photo fakery. Brassey's Inc.; 1999.

Chen L-Q, Xie X, Fan X, Ma W-Y, Zhang H-J, Zhou H-Q. A visual attention model for adapting images on small displays. ACM Multimedia Systems Journal 2003;353–64.

Chen J, Pappas TN, Mojsilovic A, Rogowitz BE. Adaptive perceptual color-texture image segmentation. IEEE Transactions on Image Processing 2005;14:1524–36.

Cohn AG. Calculi for qualitative spatial reasoning. In: Calmet J, Campbell J, Pfalzgraf J, editors. Artificial intelligence and symbolic mathematical computation. Berlin: Springer-Verlag; 1996. p. 124–43.

Ferrill E, Moyer M. A survey of digital watermarking; 1999.

Forbus KD, Birnbaum L, Wagner E, Baker J, Witbrock M, Combining analogy, intelligent information retrieval, and knowledge integration for analysis: a preliminary report. In: Proceedings of the 2005 international conference on intelligence analysis; May 2005.

Fridrich J, Soukal D, Lukas J. Detection of copy–move forgery in digital images. Digital Forensic Research Workshop; 2003.

Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis, vol. 20; 1998. p. 1254–59.

Johnson MK, Farid H. Exposing digital forgeries by detecting inconsistencies in lighting. ACM Multimedia and Security Workshop; 2005.

Lenat D. Cyc: a large-scale investment in knowledge infrastructure. Communications of the ACM 1995;38(11).

Lieberman H, Liu H. Adaptive linking between text and photos using common sense reasoning. Adaptive Hypermedia and Adaptive Web-Based Systems LNCS 2002;2347:2–11.

Meer P, Georgescu B. Edge detection with embedded confidence. IEEE Transactions on Pattern Analysis and Machine Intelligence 2001;23(12).

Ng T-T, Chang S-F, Sun Q. Blind detection of photomontage using higher order statistics. IEEE International Symposium on Circuits and Systems; 2004.

Niblack W, Barber R, Equitz W, Flickner M, Glasman EH, Petkovic D, et al. The qbic project: querying images by content, using color, texture, and shape, vol. 1908. SPIE; 1993. p. 173–87.

Perez P, Gangnet M, Blake A. Poisson image editing. ACM Transactions on Graphics 2003;22(3):313–8.

Popescu AC, Farid H. Exposing digital forgeries by detecting duplicated image regions (Technical report TR2004-515). Department of Computer Science, Dartmouth College; 2004.

Popescu AC, Farid H. Exposing digital forgeries by detecting traces of re-sampling. IEEE Transactions on Signal Processing 2005; 53(2).

Popescu AC. Statistical tools for digital image forensics. PhD thesis, Department of Computer Science, Dartmouth College; 2005.

Singh P, Lin T, Mueller ET, Lim G, Perkins T, Zhu WL. Open mind common sense: knowledge acquisition from the general public. In: Proceedings of the first international conference on ontologies, databases, and applications of semantics for large scale information systems; 2002.

Suh B, Ling H, Bederson BB, Jacobs DW. Automatic thumbnail cropping and its effectiveness. In: User interface software and technology. ACM; 2003. p. 11–99.

Swain M, Ballard D. Color indexing. International Journal on Computer Vision 1991;7(1):11–32.

Tomai E, Forbus KD, Usher J. Qualitative spatial reasoning for geometric analogies. In: Proceedings of the 18th International Qualitative Reasoning Workshop; 2004.

Wade N. It may look authentic; here's how to tell it isn't. The New York Times; Jan 24, 2006.

Sangwon Lee is a Ph.D. student at Electrical Engineering and Computer Science Department, Northwestern University since 2003. He received a M.S. in Computational Design from Carnegie Mellon University and B.S. in Engineering from Seoul National University, Department of architecture. His research focuses on image fakery detection, non-photorealistic rendering and architectural visualization.

David A. Shamma recently completed his Ph.D. at Northwestern University's Intelligent Information Laboratory. His research interests bridge artificial intelligence, information systems, and artistic installations. His art and technology installations have been displayed and reviewed internationally. Dr. Shamma recieved his M.S. and B.S. in computer science from the Institute for Human and Machine Cognition at The University of West Florida and is a member of the ACM.

Bruce Gooch is an Assistant Professor of Computer Science and Cognitive Science at Northwestern University. Perceptually optimized display is the focus of much of Professor Gooch's work. Current projects involve retargeting large images to the small displays sizes of cell phones and PDAs without affecting the perceived content and developing interactive illustrations as a move toward optimal communication content in data visualizations. He was the founding online editor of the ACM Transactions on Applied Perception. He is also an author of the books ''Non Photorealistic Rendering'' and ''Illustrative Visualization'' published by A.K. Peters. Dr. Gooch is a member of the ACM, IEEE, and Eurographics organizations.