3

A Gaze-Based Study for Investigating the Perception of Visual Realism in Simulated Scenes

MOHAMED ELHELW, MARIOS NICOLAOU, ADRIAN CHUNG, and GUANG-ZHONG YANG
Imperial College London
and
M. STELLA ATKINS
Simon Fraser University

Visual realism has been a major objective of computer graphics since the inception of the field. However, the perception of visual realism is not a well-understood process and is usually attributed to a combination of visual cues and image features that are difficult to define or measure. For highly complex images, the problem is even more involved. The purpose of this paper is to present a study based on eye tracking for investigating the perception of visual realism of static images with different visual qualities. The eye-fixation clusters helped to define salient image features corresponding to 3D surface details and light transfer properties that attract observers' attention. This enabled the definition and categorization of image attributes affecting the perception of photorealism. The dynamics of the visual behavior of different observer groups were examined by analyzing saccadic eye movements. We also demonstrated how the different image categories used in the experiments were perceived with varying degrees of visual realism. The results presented can be used as a basis for investigating the impact of individual image features on the perception of visual realism. This study suggests that post-recall or simple abstraction of visual experience is not accurate and the use of eye tracking provides an effective way of determining relevant features that affect visual realism, thus allowing for improved rendering techniques that target these features.

Categories and Subject Descriptors: I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism; I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis—*Tracking*

General Terms: Experimentation, Measurement, Performance

Additional Key Words and Phrases: Visual perception, visual realism, photorealistic rendering, eye tracking, human–computer interaction, simulation environment

ACM Reference Format:

Elhelw, M., Nicolaou, M., Chung, A., Yang, G.-Z., and Atkins, M. S. 2008. A gaze-based study for investigating the perception of visual realism in simulated scenes. ACM Trans. Appl. Percpt. 5, 1, Article 3 (January 2008), 20 pages. DOI = 10.1145/1279640.1279643 http://doi.acm.org/10.1145/1279640.1279643

1. INTRODUCTION

The quest for visual realism has been one of the main driving forces behind the rapid development of computer graphics. Visual realism, also referred to as photorealism, defines how similar a synthetic

Authors' address: Mohamed Elhelw, Marios Nicolaou, Adrian Chung, and Guang-Zhong Yang, Department of Computing, Visual Information Processing Group, Imperial College London, 180 Queen's Gate, London SW7, 2BZ, UK; email: me@doc.ic.ac.uk, g.z.yang@imperial.ac.uk. URL: http://vip.doc.ic.ac.uk/; M. Stella Atkins, School of Computing Science, Simon Fraser University, British Columbia, Canada.

image is to the photographic depiction of the same scene. The degree of visual realism required is usually dependent upon the nature of the application [Foley 2004] and, in many situations, photorealistic rendering is essential [Chiu and Shirley 1994]. In minimal access surgical simulation, the effectiveness of training of higher perceptual and spatial reasoning skills, e.g., visual recognition of a landmark in a cluttered environment, depends on how faithful the relevant scene features are represented [Liu et al. 2003]. Thus far, extensive research has been carried out in areas related to low-level vision, such as visual physiology and computational modeling of early vision [Bruce et al. 1996] and in high-level scene perception [Henderson and Ferreira 2004; Henderson and Hollingworth 1999, 1998]. However, the perception of photorealism remains an obscure process [Rademacher et al. 2001]. Although it is clear that visual realism is attributed to a combination of cues, including shadows, interreflections, depth perception, and surface attributes, such as color and texture [Bertin 1983], defining or separating specific image features in highly complex scenes is difficult. With the increasing demand for high-fidelity simulation environments and the diversity of computer graphics techniques, it is important to devise a systematic way of assessing observer responses to different rendering techniques.

This paper proposes the use of eye-tracking information to reveal the underlying features that impinge on the subjective judgment of visual realism in complex scenes. The work presented is focused on the development of objective measures for the perception of visual realism for computer-based surgical simulations. We will first discuss realistic image synthesis and the importance of taking the human visual system (HVS) into consideration for defining image fidelity metrics. In the next section, we present an experimental method for capturing the visual behavior of observers evaluating photorealism in a surgical simulation. Detailed analysis techniques and the associated quantitative results are provided in Section 3.

1.1 Photorealistic Image Synthesis

It is well recognized in computer graphics that in order to achieve realistic image synthesis, accurate local reflection and light transport models are required. Local reflection models calculate the reflected light intensity from a surface point. Existing techniques perform such computation as a function of surface orientation relative to a point light source. Surface-light interaction, which is defined by means of bidirectional reflectance distribution functions (BRDFs), must be taken into consideration. However, the limited knowledge about material reflectance properties and the trade-offs between computational costs and accuracy have led to a range of empirical and physically based local lighting techniques. A number of physically based analytic approaches have also been proposed for modeling the BRDF by computing the specular and diffuse components. Microfacet models were used by a number of researchers [Cook and Torrance 1982; Blinn 1978] to imitate specular materials and incorporate self-shadowing by using a rough surface with perfect microreflectors to account for the spectral composition of specular highlights. These models were further extended to handle a wide range of surface types and anisotropic distributions with multiple scattering from complex microscale geometries [Watt and Watt 1992; He et al. 1991]. Hanrahan and Krueger [1993] introduced a physically based technique by using Monte Carlo methods. It computes relative contributions from surface reflection and subsurface scattering and transmission, and allows for simulating layered materials and biological tissues, such as skin and green leaves. An alternative approach for BRDF modeling is to acquire measurements of the BRDF by using a goniospectro reflectometer [Matusik et al. 2003; Dana et al. 1999; Ward 1992]. However, fully representing BRDFs in a tabular form imposes prohibitive storage needs, making them difficult to utilize in a scene with many types of materials [Lawrence et al. 2004 ; McCool et al. 2001].

For light transport simulation, two methods are commonly used in the computer graphics community: ray tracing [Whitted 1980] and radiosity [Goral et al. 1984]. Ray tracing is a view-dependent method that handles only specular reflections and does not consider diffuse–diffuse or diffuse–specular interactions. Radiosity, on the other hand, provides a view-independent solution to diffuse interactions within a closed environment by subdividing the scene into discrete elements. In general, global illumination cannot be computed with a single transport mechanism and several factors have to be combined to get a realistic solution [Watt and Watt 1992]. A two-stage approach for integrating ray tracing and radiosity was proposed by Wallace et al. [1987], and was further extended to include shape-varying specular surfaces and refractions by Sillion and Puech [1989]. Most of these algorithms make certain simplifications to the BRDF related to visibility computation and the solution of the integral equation over incoming and outgoing directions [Cornell 2006].

One of the major issues for realistic rendering is to provide convincing complexity that is equivalent to that of real-world scenes [Chiu and Shirley 1994]. Generally, detailed geometry is needed for all scene objects in order to attain a realistic appearance. However, a geometry-only approach imposes enormous modeling efforts and, for complex scenes, this can result in a large number of primitives that is prohibitive for interactive rendering. Different mapping methods have been introduced to minimize the modeling complexity, which include texture mapping [Catmull 1974], bump mapping [Blinn 1978], procedural textures [Ebert et al. 1998], and level-of-detail (LOD) methods [Heckbert and Garland 1994]. Recently, image-based rendering (IBR) has established itself as a powerful alternative to geometry-based computer graphics [McMillan 1997; Gortler et al. 1996; Levoy and Hanrahan 1996]. With this technique, a set of images or depth-enhanced images is used to synthesize novel views of either synthetic or real environments. Depth information is used to project source image pixels into a 3D space and then reproject them onto a target image plane. The main advantage of image-based rendering is increased rendering realism by accounting for global illumination and interobject reflections. Other gains include efficient modeling, representation and rendering of complex, objects, and enhanced overall system performance. However, for multiple source images, issues related to visibility and reconstructions are still to be resolved. Moreover, a large number of images may be required to fully model the scene, thus resulting in huge storage requirements.

1.2 Human Perception and Realistic Graphics

Realistic image synthesis aims to attain high fidelity visual appearance as perceived by humans. It takes into account observer characteristics, in addition to physical modeling in image rendering [Larson et al. 1997; Ferwerda et al. 1996; Rushmeier et al. 1995; Ward 1994; Tumblin and Rushmeier 1993]. Thus far, several image metrics have been introduced for assessing the perceptual quality of different rendering techniques by imitating the behavior of the HVS. In general, three main sensitivities to variations of the HVS are considered by most image-quality metrics: relative luminance variations, nonlinear eye behavior as a function of luminance, and spatial frequency of luminance variations [McNamara 2001], Daly [1993] presented the visible difference predictor (VDP) that uses features of the HVS to produce maps of local probabilities of perceivable differences between images. Lubin [1995] proposed the use of the Sarnoff visual discrimination model (VDM), which takes as input two images, and produces as output a map of just-noticeable differences (JND). The VDM puts more emphasis on modeling the physiology of the visual pathway and, hence, operates in the spatial domain as opposed to the frequency domain considered by the VDP [McNamara 2001].

The established quality metrics such as the VDP and VDM, however, experience a number of limitations [Li et al. 1998]. For example, these models consider only certain aspects of the complex HVS. Furthermore, they can have high computational or storage requirements and may not be well suited for interactive 3D applications. Despite these limitations, image fidelity metrics have been used to guide the image-generation process. Computational resources can be optimized toward image regions capturing most of visual attention and thus high-quality rendering can be achieved while efficiency is maintained. Mitchell [1987] used visual models to determine image sampling rates and produce antialiased images. A frequency-based ray tracer was introduced by Bolin and Mayer [1995], which used a perceptual model to detect visibility artifacts and control the sampling process. HVS models were also used in global illumination solutions to effectively control the progression of the rendering procedure [Bolin and Meyer 1999; Myszkowski 1998].

Considering the fact that humans will observe the images at the end, several image-quality metrics and assessment approaches include decisions made by human observers. For example, a number of psychophysical experiments have been proposed for evaluating the quality of rendering methods and verifying some of the introduced image-quality metrics. McNamara et al. [2000] conducted experiments involving human judgment of lightness in order to evaluate the fidelity of different rendering techniques. Yee et al. [2001] proposed the use of spatiotemporal error tolerance maps in order to accelerate global illumination computation for animations. Other work included the use of human visual perception for selective rendering and illumination of complex scenes [Dumont et al. 2005; Sundstedt et al. 2005; Stokes et al. 2004] and the investigation of feature saliency [Itti et al. 1998] during evaluation of visual realism [Rademacher et al. 2001].

Thus far, most published psychophysical experiments were carried out under simplified conditions, e.g., stimuli comprised of basic geometrical shapes, such as, cubes and spheres with varying visual factors, such as, shadow softness, surface smoothness, and lighting conditions. In practice, the generalization of these results to determine their effects in complex computer-generated images is a complicated problem, since there are significant interactions between the different dimensions of the visual experience [Greenberg 1999]. Moreover, questionnaire-based evaluation approaches used for assembling and assessing users' judgment of images suffer from inherent limitations. For example, conscious responses give information at a high level of processing without exactly pointing at image features considered during visual examination. Furthermore, subjective evaluations can be inconsistent [Bertrand and Mullainathan 2000] as they are influenced by factors, such as observers' understanding of visual realism and in defining rating scale endpoints. There are also potential problems with post-test evaluation, because of inaccurate recall and memory effects, such as recency [Aldridge 1995]. Finally, the instructions given during the task and the training methods used may also influence scores [IJsselsteijn et al. 2000; Freeman et al. 1999].

Although extensive research has been carried out in the areas of visual perception, realistic rendering, and perceptually-based rendering, the problem of determining what in the image makes it appear photographic or synthetic remains largely unexplored [Rademacher et al. 2001]. In this paper, we propose the use of eye-tracking information to resolve or substantiate image features that subliminally determine the perception of photorealism and identify salient scene regions that draw most of observers' attention.

1.3 Eye Tracking and Visual Attention

Eye tracking has been used since the 1960s as an objective tool for the analysis of visual attention and perception, although the early technology was difficult to use, invasive, and not widely adopted. The essence of eye tracking is to measure where the gaze is directed and it relies on the fact that the human visual system provides the best visual acuity at the fovea that has a small visual angle of around 2° [Yang et al. 2002]. The acuity drops off rapidly beyond the fovea toward the periphery. Therefore, the eyes have to scan the scene with a series of fixations in order to build up a detailed map of the scene. The rapid movements between these fixations are called saccades. Fixations are linked to the user's conscious attention [Kundel 2004; Velichkovsky et al. 2000] and there are many studies that employ eye tracking to determine attention in applications, such as reading [Rayner 1998], visual search [Pomplun et al. 2001], and image understanding [Dempere-Marco et al. 2002]. Kundel et al.

[1978] developed a model for using clusters of fixations for analyzing scan paths during visual search, where it was assumed that a fixation picks up a certain amount of visual information from both foveal and peripheral inputs. A fixation cluster is a grouping of fixations within close spatial proximity and it takes into account the fact that when an observer focuses attention on image features, multiple fixations sometimes occur. Eye-tracking studies have been used in radiology training for many years, where the duration of fixations on areas with lesions has been used to classify certain types of diagnostic errors, e.g., to determine whether missed lesions were recognized as being anomalous or whether there was a decision error [Atkins et al. 2005; Kundel 2004; Nodine and Kundel 1987; Kundel et al. 1978]. For laparoscopic surgery training, fixation information has been used in conjunction with assessing eye-hand coordination [Nicolaou et al. 2004; Law et al. 2003]. In recent years, gaze-contingent systems have been introduced in areas such as level-of-detail rendering, determining saliency of 3D polygonal models, collision handling, multirsolution displays, and rendering of virtual environments [Howlett et al. 2004; O'Sullivan et al. 2003; Murphy and Duchowski 2001; Loschky and McConkie 2000; Ohshima et al. 1996]. These are applications where the information presented to the viewer is manipulated to match human-processing capability, often with foveal and peripheral perception matching in real time. A variant of gaze-controlled rendering was proposed after the introduction of inattentional blindness [Mack and Rock 1998], where unattended stimuli are not perceived by observers, and, hence, the term blindness. In this case, human observers' failure to notice degradations in the quality of image details unrelated to their assigned task allows for efficient rendering systems without loss of perceived visual quality [Cater et al. 2003; Rensink 2000].

Despite the amount of work on eye tracking and scene analysis, limited effort has been directed towards using eye-tracking information for studying scene features affecting the perception of visual realism. In a survey of eye-tracking applications, Duchowski [2002] showed that a major difficulty in extracting useful information about scene perception arises because any task given to observers might bias their visual behavior when they view a scene. In fact, the viewing strategy and eye-movement patterns change as a function of the viewing task [Yarbus 1967]. Duchowski [2002] proposed that one way to address this issue is to give observers a task that does not force the creation of a coherent memory representation for the scene. In this work, we performed eye tracking on observers carrying out the task of evaluating visual realism of real and synthetic images. Such task should not bias observers' visual behavior and, by analyzing where the gaze is directed, it is possible to investigate eye-movement behavior in complex scenes, such as those found in surgical simulation. Thus it is possible to determine which scene aspects, i.e., features, attract observers' attention for the given task. It should be noted that no attempt was made to simplify the study by using reduced stimuli, such as simple or abstract geometric shapes, or by manipulating individual scene features, e.g., shadows and edge smoothness, as done in other psychophysical studies investigating the perception of photorealism [Rademacher et al. 2001; McNamara et al. 2000]. This is because of the fact that different scene features are not mutually exclusive [Greenberg 1999] and the full stimuli has to be entirely presented for faithful investigation to be achieved.

1.4 The Scope of the Paper

This paper presents a study where eye tracking was used to identify salient image features affecting the perception of visual realism. The methodology used compares rendered images generated with varying visual effects versus actual *in vivo* images. The variability of features thought to be important to visual perception depends on the rendering technique itself and, in this study, they include texture quality and specular highlights. There are two parts to the method. In the first part, subjective evaluation of the realism of single images is made using a five-point Likert scale. The users' realism ratings are used to substantiate the importance of varied computer graphics (CG) elements. Eye-tracking information corroborates fixation of those elements, since fixations provide a higher bandwidth of information than a single-valued result from a questionnaire. In the second part, pairwise image comparisons are made. The scores are statistically analyzed via Wilcoxon signed-ranks test and post-hoc pairwise comparisons, and correlated with the fixations on specific image regions. We specified simple tasks to determine how realistic single images are in the first experiment and which of two images appears to be the most realistic in the second experiment. Indeed, if we had used a high-level task, such as determining whether an image contained a lesion, the observers would likely attend to different features in the image. The method used for this study is independent of the rendering technique used to acquire the synthetic images and can be generalized to other types of images, such as those of natural scenes, hence, providing the basis for a general framework for evaluating visual realism in computer graphics.

2. METHOD

2.1 Stimuli Generation

We used clinical bronchoscopy images for stimuli. Bronchoscopy is a diagnostic procedure in which a tube with a tiny camera on the end (an endoscope) is inserted through the nose or mouth into the lungs. The procedure provides a view of the airways of the lung and allows doctors to collect lung secretions or tissue specimens (biopsy). Bronchoscopy is a difficult procedure to perform and there are significant clinical benefits that can be obtained by using computer-based training simulators [Ost et al. 2001]. The stimuli used in this study included both real and synthetic images where the latter were generated with an image-based rendering method for visualizing the bronchial lumen. The method relied on deriving matched shading and texture parameters directly from video bronchoscope images [Chung et al. 2004]. With this approach, 2D/3D registration was first used to match video bronchoscope images with a 3D polygonal model of the bronchial tree derived from a CT scan. This enabled the exact camera pose of the bronchoscope examinations to be identified.

Surface details, including texture and shading parameters were also extracted. The texture map was derived directly from the video bronchoscope images. The shading parameters were recovered by modeling the bidirectional reflectance distribution function ρ of the visible surfaces, based on assumptions of tissue uniformity by using a cubic curve parameterized on γ, the cosine of the angle between the viewing vector V and surface normal N [Chung et al. 2004]:

$$\rho(V, N) = \sum_{0}^{3} c_i B_i^3(\gamma) \tag{1}$$

where

$$\gamma = \frac{V.N}{|V|\,|N|}, B_i^n(t) = \frac{n!}{i!(n-i!)}(1-t)^{(n-i)}t^i \tag{2}$$

The estimated BRDF was then used to predict the expected shading intensity, such that a texture map, independent of lighting conditions, could be factored out of each video image and back-projected onto the bronchial model. The derived texture maps were then merged to form a single texture atlas which, combined with the BRDF model, allowed for rendering new views not encountered in the original video bronchoscopy with a high degree of photorealism. For further enhancement of photorealism, the specular properties of the lumen were modeled by using computer-generated noise [ElHelw et al. 2004], which allowed for rendering realistic highlights. 2D Perlin noise [Perlin 1985] was first used to create a noise image, i.e., image-storing noise samples, which was then converted to a reflectance map. For each point in the noise image, the associated tangent vectors to the surface were defined and each pixel in the reflectance map encoded the direction that the corresponding point was facing by using three channels of information, which were conveniently mapped to a standard RGB image.

During runtime, per-triangle reflectance information was extracted by means of texture mapping for calculating specular highlights. The extracted vectors were defined in image coordinate space, so they had to be transformed into a coordinate space that is local to the triangle being processed. The basis vectors for such coordinate system, known as the surface-local coordinate space, can be defined by using the surface tangent T, bitangent B, and normal N. The first two vectors were computed from the partial derivatives of the triangle object space coordinates with respect to its corresponding texture coordinates [Fernando and Kilgard 2003]:

$$T = \left(\frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}, \frac{\partial z}{\partial u}\right) = \left(-\frac{B_o}{A_0}, -\frac{B_1}{A_1}, -\frac{B_2}{A_2}\right) \tag{3}$$

and

$$B = \left(\frac{\partial x}{\partial v}, \frac{\partial y}{\partial v}, \frac{\partial z}{\partial v}\right) = \left(-\frac{C_o}{A_0}, -\frac{C_1}{A_1}, -\frac{C_2}{A_2}\right) \tag{4}$$

where

$$\begin{aligned}
(A_0, B_o, C_o) &= [(x_1, u_1, v_1) - (x_0, u_0, v_0)] \otimes [(x_2, u_2, v_2) - (x_0, u_0, v_0)] \\
(A_1, B_1, C_1) &= [(y_1, u_1, v_1) - (y_0, u_0, v_0)] \otimes [(y_2, u_2, v_2) - (y_0, u_0, v_0)] \\
(A_2, B_2, C_2) &= [(z_1, u_1, v_1) - (z_0, u_0, v_0)] \otimes [(z_2, u_2, v_2) - (z_0, u_0, v_0)]
\end{aligned}$$

In the above equations $(x_0, y_0, z_0), (x_1, y_1, z_1), (x_2, y_2, z_2)$ *and* $(u_0, v_0), (u_1, v_1), (u_2, v_2)$ represent the prospective triangle's object and texture spaces coordinates, respectively, and $\otimes$ denotes the cross product operation. Subsequently, N was calculated from the cross product of T and B or alternatively the normal supplied with the original model could be used. After computing the basis vectors of the surface-local coordinate space, the graphics processor unit (GPU) was used to efficiently transform the extracted image-space reflectance vectors to the surface-local space and carry out specular highlight computations.

2.2 Types of Stimuli and Controlled-Image Factors

Similar views were used for both real and synthetic stimuli in order to investigate the observers' visual behavior when viewing real and rendered scenes. The stimuli show static bronchoscopy images corresponding to views of the bronchial tree from three different camera poses labeled A, B, and C, as shown in Figure 1. For each pose, five images were presented comprising four rendered images plus a real image, referred to as image categories 1–5, respectively. Therefore, the total number of unique stimuli images is 15. Examples images of categories 1–5, for one of the poses are shown in Figure 2. Image categories 3–1 differ in the texture detail dimension where categories 2 and 1 were acquired by gradually filtering (Gaussian smoothing kernels with spatial support 3×3 and 7×7, respectively) the original texture image (category 3 image). Category 4 is identical to category 3 with added specular highlights, whereas category 5 represents the real image. The five categories 1–5 can be ordered with an ascending visual quality.

2.3 Experiment Setup

2.3.1 *Participants.* The same sixteen observers took part in the two experiments, which were carried out sequentially. All observers had normal or corrected to normal vision. Observers were divided into two groups: 5 observers with extensive graphics and tissue appearance knowledge, 1 of whom was female, and 11 observers without such knowledge, 6 of whom were females.

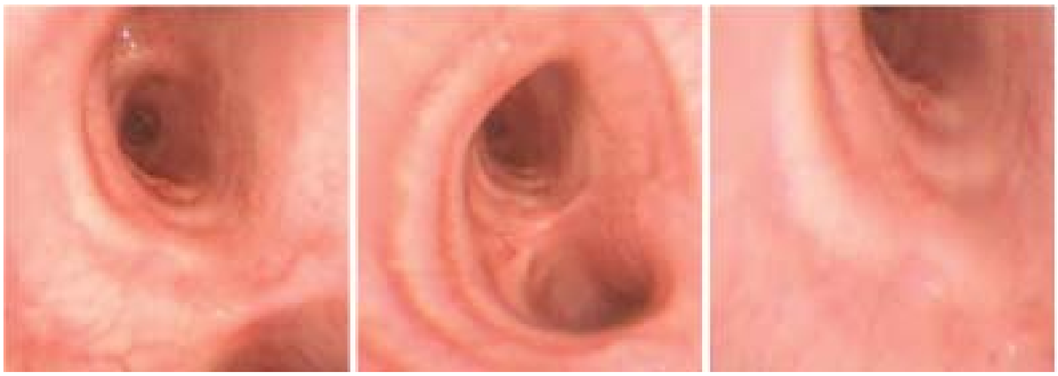


Fig. 1. The stimuli show three views of the bronchial tree corresponding to camera poses A, B, and C (left to right).

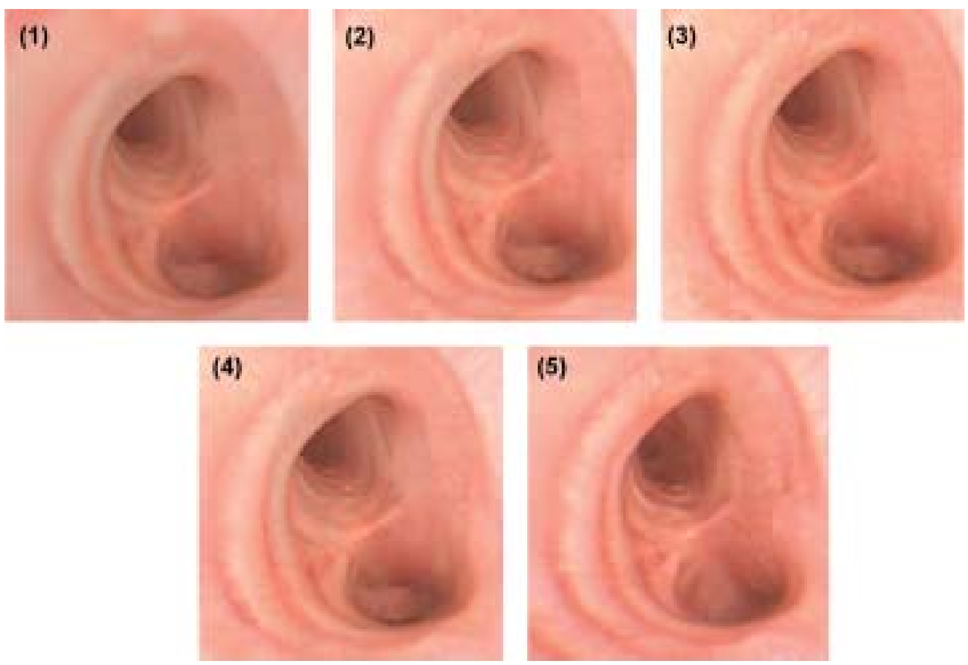


Fig. 2. For each camera pose, five images representing categories 1–5 are used for the stimuli. Shown above are images for pose B. Categories 2 and 1 were acquired by gradually filtering (Gaussian smoothing kernels with spatial support 3×3 and 7×7, respectively) the original texture image (category 3 image). Category 4 is identical to category 3 with added specular highlights, whereas category 5 represents the real captured video bronchoscope image.

2.3.2 *Equipment.* The participants were standing 60–70 cm from the 17″ display. The images were shown with a resolution of 800×600 pixels. The illumination conditions in the room were constant throughout the experiments. Observer's eye movements were measured by using a Tobii ET-1750 eye tracker [Tobii 2006], which is an infrared video-based binocular eye-tracking system recording the position of the fixations on the screen displaying the stimuli, at about 30 samples per second. A fixed infrared light source is beamed toward the eye, while a camera below the monitor records the position of the reflection (known as the Purkinje reflection) on the cornea surface relative to the pupil center.

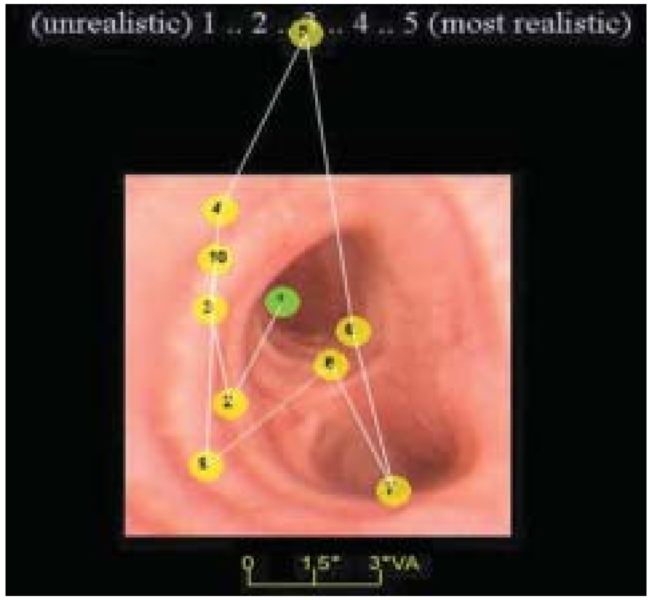


Fig. 3. A sample screen shot from Experiment 1 with fixations shown as numbered nodes in sequential order. The lines represent the scan path between successive fixations. The evaluation scale is displayed at the top center of the screen. The visual angle scale is shown at the bottom of the image.

The infrared images of the eye position were digitized and processed in real-time. Each observer's eye movements were first calibrated, by fixating on up to five points (six samples per point) on the monitor until the residual error of alignment of the points had been minimized. After the calibration procedure, which took about 20 seconds, the point of regard on the display screen was determined with an average accuracy of 0.5° visual angle, i.e., error of +/− 0.5°, across the screen [Tobii 2006]. A certain amount of head movement was allowed; good motion compensation combined with binocular eye tracking resulted in minimal data loss (less than 4%).

2.3.3 *Experiment 1: Single-Image Evaluation.* The observers were first shown an example image each of category 1 (most unrealistic) and category 5 (real) displayed side-by-side for mental calibration. Then they were presented with a series of 15 images (the five renderings on each of the three poses), and asked to rank each image in terms of visual realism using a Likert scale (1–5), with no timing limits. Images were displayed in random order, the same for each observer, and observers were not told which the real images were, nor were they informed how the images were obtained. The evaluation scale (without further example images) was displayed at the top of the screen for reference, as illustrated in Figure 3. It is worth noting that the image in the slide is of a size that subtends the same visual angle (8.2°× 9.5°) as the one displayed on the larger, but more distant monitor during a real surgery. A researcher sitting nearby controlled the presentation of stimuli on the screen and recorded the score for the stimuli, which were spoken aloud by the observer.

2.3.4 *Experiment 2: Side-by-Side Image Comparisons.* In Experiment 2, a two alternative forced-choice (2AFC) task was conducted using images only from categories 3, 4, and 5. We did not use images from categories 1 and 2, because these were too obviously identified as unreal. Such decision was further supported by the results of Experiment 1. The observers viewed 18 side-by-side image pairs and were asked to choose which of the two (left or right) was the most realistic. The image pairs were always from different categories and poses. Different poses were used, because a pilot study showed it was too easy to spot the differences when comparing similar poses.

2.4 Statistical Analysis

Statistical analysis was carried out using SPSS software [SPSS 2006]. All the tests used were two-sided with level of significance set to 0.05. Data was tested for normality using Kolmogorov–Smirnov test [Sidak et al. 1999] and it was found to be not normally distributed. Nonparametric tests were subsequently used for further analysis. First, the Friedman test was used to assess an overall inter-category variation. In order to isolate differences between image categories, post-hoc multiple comparisons for paired samples (Wilcoxon signed-ranks test) were then performed for all possible combinations between categories considering as statistically significant a p value of 0.005 (calculated as 0.05/number of comparisons = 10).

2.5 Eye-Movement Analysis

In order to capture eye-movement dynamics in the side-by-side experiment, a discrete-time Markov chains approach was used. Markov modeling has been used in other work to investigate the sequence of temporal fixations [Dempere-Marco et al. 2002; Hacisalihzade et al. 1992]. In this paper, we focus on a discrete time, discrete state space Markov chain, based on a first-order Markov process, to model eye-movement dynamics. The sequence of fixations can be described with a stochastic process $\{X_n\}$ in discrete time $n = \{0, 1, 2, \dots\}$ over random image locations; thus discrete state space S. Such a process is called a Markov chain, if for all times $n \geq 0$ and all states $i_0, i_1, \dots, i, j \in S$

$$P\left(X_{n+1} = j \,|\, X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0\right) = P\left(X_{n+1} = j \,|\, X_n = i\right) = P_{ij} \tag{5}$$

where P_{ij} denotes the probability that the chain in state i moves into state j (one-step transition probability). To determine the transition probability P_{ij}, between the states i and $j \in S$, the number of transitions t, per state were calculated for all the states. For example, in case of the five image features (F1–F5) defined in Section 3.1, each feature can be designated as one state for the Markov model leading to a finite state space S = F1, ..., F5. The inter- and intrastate transition probabilities are given by Nicolaou et al. [2004]:

$$\begin{pmatrix} t_{11} & \cdots\cdots\cdots & t_{15} \\ \vdots & . \quad . \quad . & \vdots \\ \vdots & . \quad . \quad . & \vdots \\ \vdots & . \quad . \quad . & \vdots \\ t_{51} & \cdots\cdots\cdots & t_{55} \end{pmatrix} = \begin{pmatrix} p_{11} = \frac{t_{11}}{t_{11}+t_{12}+t_{13}+t_{14}+t_{15}} & \cdots\cdots\cdots & p_{15} = \frac{t_{15}}{t_{11}+t_{12}+t_{13}+t_{14}+t_{15}} \\ \vdots & . \quad . \quad . & \vdots \\ \vdots & . \quad . \quad . & \vdots \\ \vdots & . \quad . \quad . & \vdots \\ p_{51} = \frac{t_{51}}{t_{51}+t_{52}+t_{53}+t_{54}+t_{55}} & \cdots\cdots\cdots & p_{55} = \frac{t_{55}}{t_{51}+t_{52}+t_{53}+t_{54}+t_{55}} \end{pmatrix} \tag{6}$$

3. RESULTS

3.1 Experiment 1: Fixation Categorization

Visual inspection of the distribution of fixations in the single image evaluation experiment resulted in the identification of a number of different image areas that drew most of the attention. Manually labeling these common areas resulted in a set of image features into which fixation data could be grouped. A description of these features is given in Table I and enlarged examples are shown in Figure 4. It is worth noting that the centers of the features were 0.5 to 3.0 degrees apart in terms of visual angle.

Figure 5 shows the percentage of normalized dwell time per feature for each observer [Hu et al. 2003]. Dwell times were normalized by the area of each visual feature within the image. Observers 1–5 are the expert group and observers 6–16 are the novices.

Table I. Feature Categorizations and Codes

Feature Description	Feature Code
Light reflections/specular highlights	F1
3D surface details	F2
Depth visibility	F3
2D texture detail	F4
Edges/silhouettes	F5

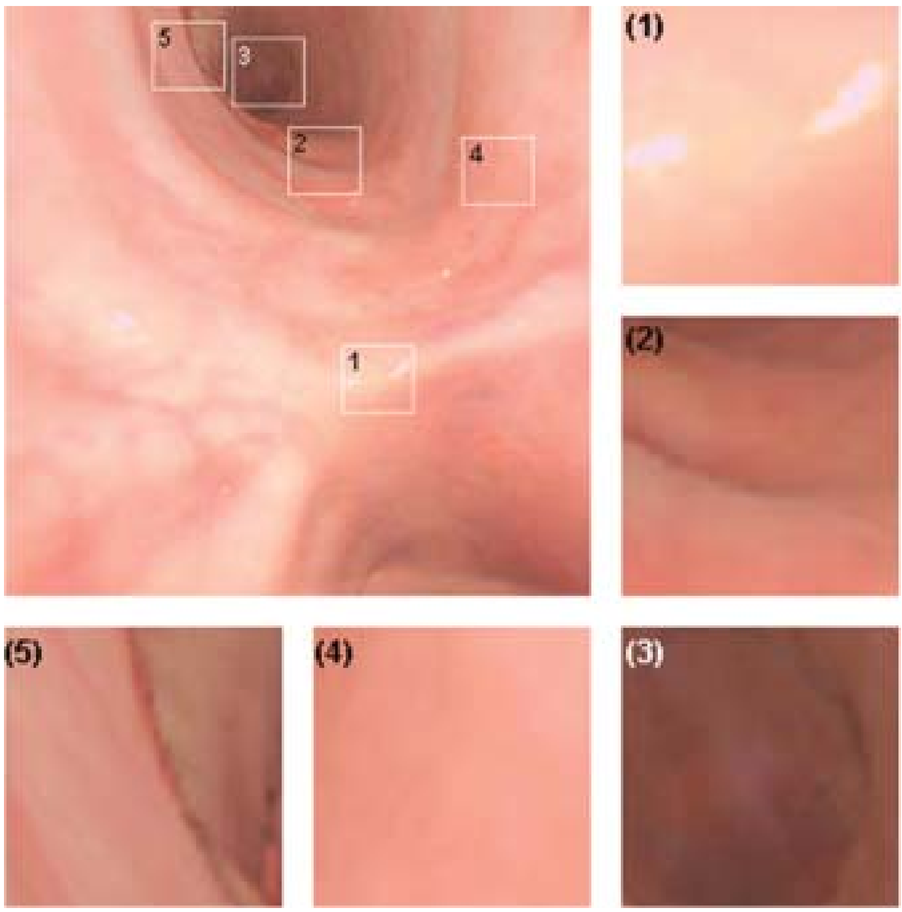


Fig. 4. An example of image features shown in blown-up views (pose C). Feature (1) corresponds to light reflections/specular highlights, feature (2) 3D surface details, feature (3) depth visibility, feature (4) 2D texture detail, and feature (5) edges/silhouettes. The aliasing artifacts found in some images are because of magnification.

It can be observed that some image features were fixated upon more than the others, when evaluating photorealism. For instance, F1, F2, and F3 appear to be the most observed features. However, all the observers do converge to essentially inspecting the same features. Figure 6 further illustrates the distribution of the normalized dwell time per feature for all observers. It is worth noting that there were no significant differences in the percentage of dwell time per feature between the novice and expert groups.

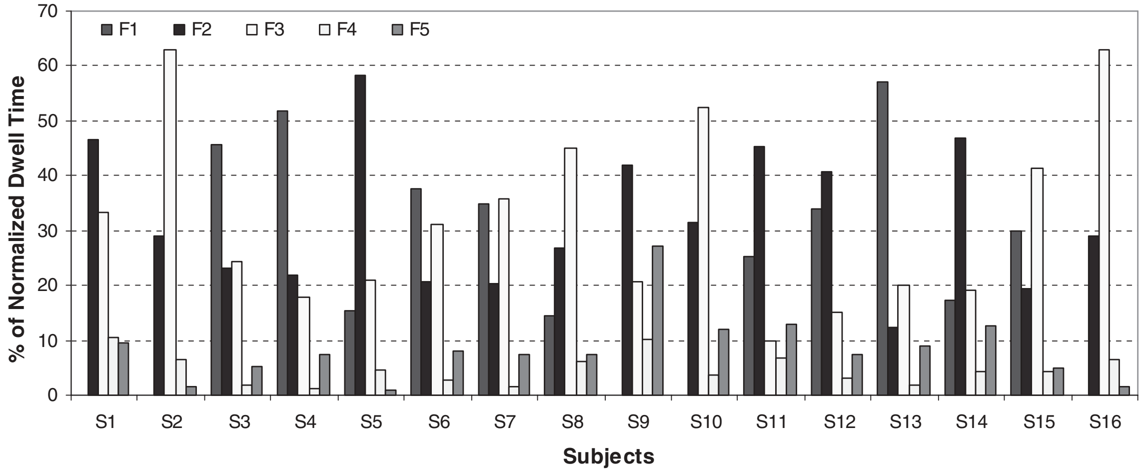


Fig. 5. The percentage of normalized dwell time per feature for each observer.

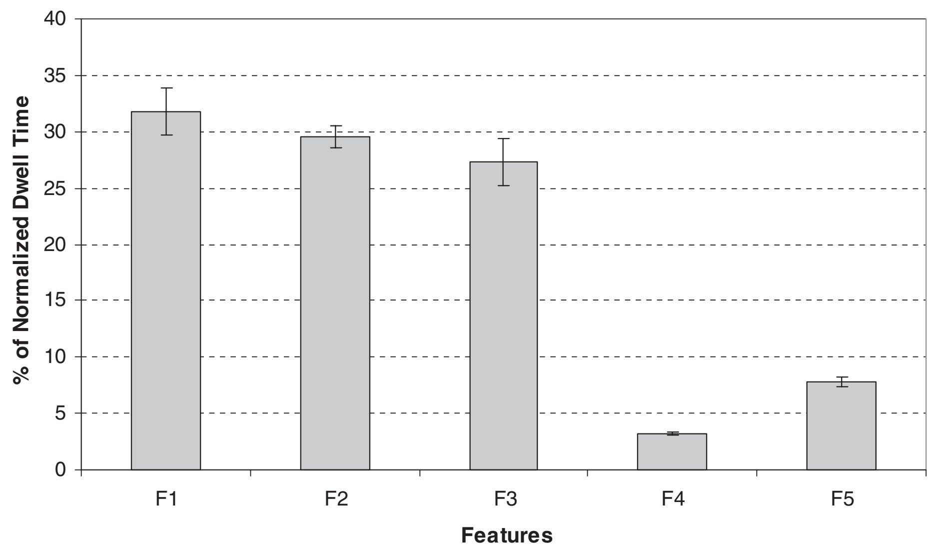


Fig. 6. The percentage of normalized dwell time per feature for all observers. Error bars show 1 standard deviation.

3.2 Experiment 1: Evaluation of Realism Scores

The scoring results for each observer were recorded and analyzed following the procedures, as outlined in Section 2.4. We hypothesized that the scoring for images from different categories should be different, corresponding to the difference in perceived visual realism. Figure 7 illustrates the average score given for each image category. It can be seen that the five categories scored differently. Statistical analysis demonstrated significant differences between image categories (Friedman test $\chi^2 = 41.962$, $df = 4$, $p < 0.001$). Additional intercategory comparison analysis demonstrated significant differences between all categories when compared to category 5 (the real image), except when compared to categories 4 and 3, as shown in Table II, suggesting that category 4 images were perceived to be close to photorealistic.

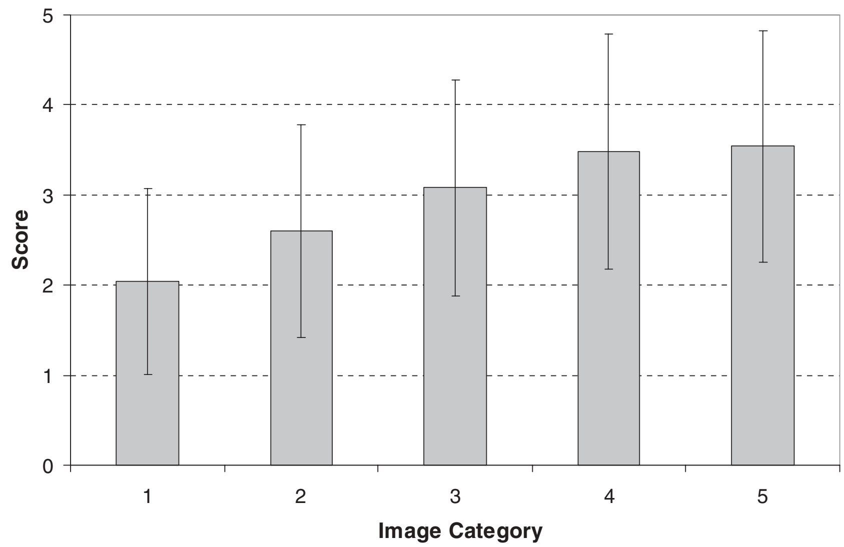


Fig. 7. Mean recorded score for each image category. It can be seen that the five rendering categories from 1 to 5 were perceived with an increasing amount of visual realism (error bars indicate the scale of 1 standard deviation).

Table II. Multiple Comparisons Analysis for all Image Categories Compared to Category 5 (Real Image)[a]

Comparison	z	2 Tailed Asymptotic p-value	Significant
Category 5 versus category 1	−4.393	0.001*	Yes
Category 5 versus category 2	−3.132	0.002*	Yes
Category 5 versus category 3	−2.092	0.036	No
Category 5 versus category 4	−0.251	0.802	No

[a](*Wilcoxon signed-ranks test with $p \leq 0.005$ was considered significant).

These results demonstrate that a salient image feature, in this case, specular highlights, could be linked to the increased perception of visual realism.

3.3 Experiment 2: Image Comparisons and Dynamic Eye-Movement Analysis

In the 2AFC test, the result of comparing Categories 3 and 4 images side-by-side in terms of realism showed that category 4 was chosen as more realistic than category 3, 61.5% of the time. Evidently the missing specular highlights in category 3 reduced the observers' perception of visual realism. The result of comparing category 3 with category 5 images showed the latter was selected as being more realistic 64% of the time. However, when comparing category 4 with category 5 images, the real images were selected as being more realistic only 39% of the time. These results corroborate the findings of Experiment 1, where Categories 4 and 5 images were perceived with similar visual realism. They also demonstrate how the addition of a salient image feature, in this case, specular highlights, positively affected observers' judgment of realism.

Insights into observers' search strategies can be gained by inspecting their eye-movement patterns when observing side-by-side image pairs. The image pairs that had received approximately equal ratios of forced choices for each of the component images were considered to be "similar" and used for further analysis. Figure 8 shows the captured eye movement behavior of two observers, one expert and one naïve while observing a similar image pair. Qualitative inspection of the expert's fixation transitions reveals a consistent search pattern saccading between similar intra- and interimage features, as well

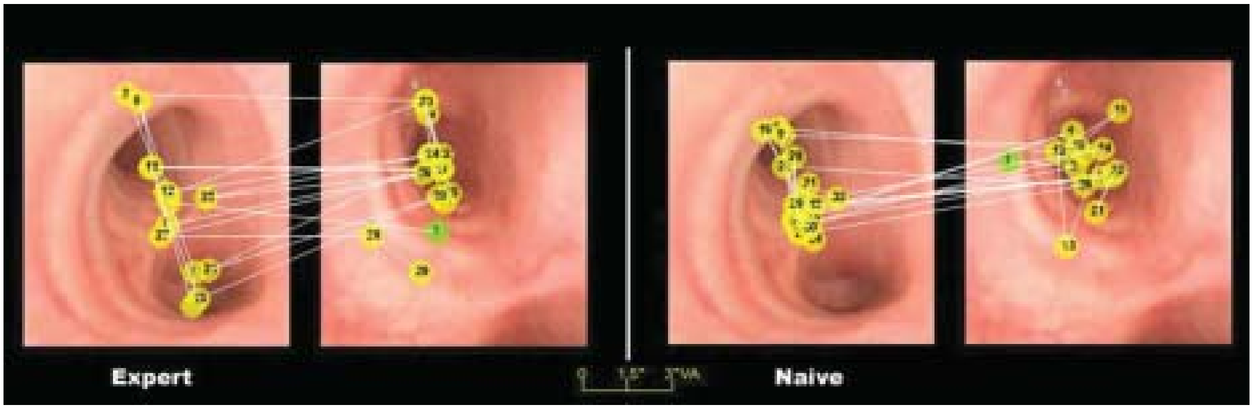


Fig. 8. Eye movement behavior for an expert (left) and naïve (right) observers looking at the similar image pair. Notice the different visual behavior between the two observers.

Table III. Normalised Fixation Distribution by Feature Area for the Similar Image Pair in Figure 8 Across All Observers[a]

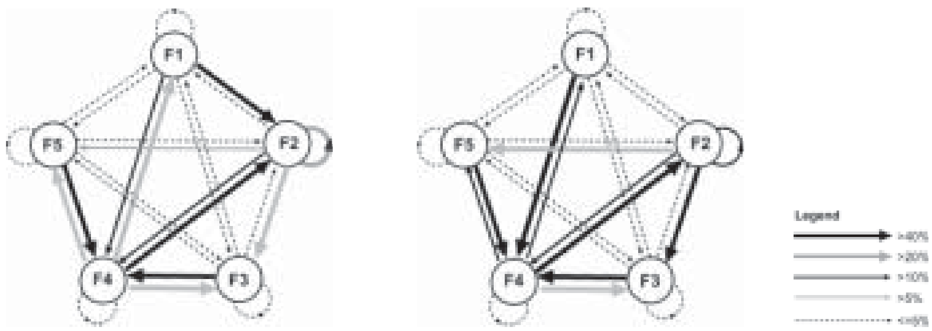
Feature	LL	RL	LR	RR
F1	65.3%	11.7%	26.9%	0%
F2	20.2%	35.8%	60.9%	11.8%
F3	9.3%	44.8%	5.6%	69.3%
F4	2.3%	3.2%	2.5%	3.4%
F5	2.9%	4.5%	4%	15.5%

[a] LL (LR) = Looking at left image choosing left (right) as being more real, RL (RR) = Looking at right image choosing left (right) as being more real.

as paying attention to peripheral features. The naïve observer, however, shows a lack of a well-defined strategy and concentrating more on central features.

Table III demonstrates the distribution of fixations by feature (using the total dwell time for each feature normalized by feature abundance [Hu et al. 2003]) for all observers while looking at each component image and taking into account their photorealistic preference. The results derived correspond well to those of Experiment 1, highlighting the importance of F1 to F3 in resolving photorealistic details. It is also interesting to note the varying feature importance when viewing layout changes, i.e., two images per slide. F1 (specular highlights) seems to be the dominant feature visualized on the left component image for observers who thought it was the most photorealistic of the two. Conversely, for observers who thought that the right image was more real, the dominant feature visualized was F2 (3D surface details).

In order to quantify the differences in visual search behavior and reveal the feature states that formed the underlying viewing strategy, a Markov transition matrix normalized by feature area was computed [Dempere-Marco 2004]. Features F1–F5 were designated as the states for the Markov model, as described in Section 2.5. Figure 9 illustrates the transition matrices for the same expert and naïve observers from Figure 8, while visualizing all 18 image pairs. For each feature, the probability of transiting to the same or other features is denoted by the thickness of the arrows coming out of each node. This reveals the weight in terms of transition probability with which different features are compared and the contrastingly different behavior between the expert and naïve observers. This figure also shows that



(a) Expert Subject (b) Naïve Subject

Fig. 9. Transition matrices for normalized fixation probabilities between features 1–5 for all pairwise images for (a) an expert and (b) a naïve observers. Self-pointing lines represent intrafeature transitions. The thickness of the arrows corresponds to the interstate transition probability between image features and, hence, their significance when evaluating photorealism.

although the relative saliency (after normalization with feature abundance) of texture details (F4) is low, it is used by both observers to ascertain the appearance of specular highlights and 3D surface/depth details. For the naïve observer, however, it seems that there is significant effort in differentiating between surface texture details and other visual features, as evident from the high percentage pairwise transition values at F4. This indirectly suggests the lack of experience of the observer and the underlying reason of the visual search patterns, as shown in Figure 8.

3.4 Subjective Responses

At the end of session, each observer answered a questionnaire to describe what image aspects they considered in their evaluation. Not surprisingly, for many users the questionnaire results were not consistent with their visual behavior. This is because of the fact that many of the visual search tasks were performed subconsciously, and postrecall of visual experience is not reliable. For example, 7/16 users denied considering edges in their evaluation of realism, but their fixation data revealed the opposite. Furthermore, 8/16 of the observers stated in the questionnaire that a feature was important or very important, however, they spent less than 50% of their fixations on that feature. Also, 5/16 observers said a feature was not of any importance, but they focused on it for more than 25% of the time. This supports the hypothesis that subjective answers to questionnaires do not always represent underlying visual attention and that eye tracking can be used to complement subjective studies.

4. DISCUSSION AND CONCLUSIONS

Objective investigation of salient image features determining the level of perceived visual realism is a crucial component for achieving photorealistic rendering. Current knowledge of the human visual system has resulted in a number of techniques for improving the realism of displayed images. Moreover, several image fidelity metrics have been developed for evaluating the quality of rendered images. However, only certain aspects of the complex HVS are considered by these methods. By taking human judgment into consideration, a number of psychophysics-based approaches were proposed to evaluate the realism attained by different rendering techniques and determine salient features that make an image perceived as realistic. Thus far, only simple scenes have been examined and many were under constrained conditions.

Visual perception is a complex experience and this paper describes the use of eye tracking to investigate the visual behavior of subjective inspection of complex scenes. Views from a video bronchoscopy were used as the experimental stimuli, such that for each view, images with varying visual qualities were generated. Images were presented to each observer for visual realism evaluation and two experiments were conducted. In the first experiment, observers were asked to evaluate a number of images with different visual quality presented one at a time. It was observed that all observers fixated on image areas that correspond to meaningful contextual scene details, which helped the identification of visually salient image features. In the second experiment, side-by-side image pairs were displayed and observers were asked to pick the image closer to reality. It was noticed that the inspected image features were similar to those in Experiment 1 and that different observer groups had different visual behavior. The hypotheses that the different image categories were perceived with varying degrees of visual realism and that the use of image-based rendering with detailed modeling of specular reflection can improve the quality of visual realism were statistically supported. It should be noted, however, that prior knowledge and preexisting behavior of the observers may have played a role in the visual search behavior, which cannot be completely excluded. However, the task given in the experiments, i.e., assessing visual realism rather than searching for specific lesions, has forced observers to look for features that are most relevant to visual realism.

This study suggests that if important features can be identified, then the efficiency of existing rendering techniques can be increased by allocating computational resources to features or regions of the image that are mostly noticed by the viewers [Longhurst and Chalmers 2004]. For example, from the current study, it can be seen that the viewers paid much visual attention to or around areas where specular highlights were present and that the geometric details of edges/silhouettes were the second most significant feature in determining visual realism. Once such visual factors are established, rendering methods particularly targeting these factors can be developed. This paper presents a first study aimed toward these objectives. The accuracy of the study can be further improved by using larger number of poses. Additional results related to quantifying visual search behavior of naïve and expert observers can be attained by computing the combined transition matrices for both groups. Further work investigating the relationship between individual image features and perceived realism is desirable. This needs to provide more control over the stimuli to examine if quality degradation or exclusion of some features result in different visual behavior or have an impact on the scoring of realism. Another natural extension to the proposed study is to investigate salient features in dynamic scenes where static images are replaced with video sequences. One of the major challenges for dynamic scene analysis, however, is that the scan path of the eye is influenced not only by spatial information but also by temporal factors, such as velocity-dependent contrast sensitivity [Yee et al. 2001]. The effects of these factors on higher-level perceptual tasks, such as evaluating visual realism, should be taken into account.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Thanos Athanasiou from Imperial College London for his help with the statistical analysis.

REFERENCES

Aldridge, R. 1995. Recency effect in the subjective evaluation of digitally-coded television pictures. In *Fifth IEEE International Image Processing Conference*, Edinburgh, 1995. IEEE Press, Los Alamitos. CA. 336–339.

Atkins, M. S., Moise, A., and Rohling, R. 2005. An application of eyegaze tracking for designing radiologists' workstations: Insights for comparative visual search tasks. *Trans. Appl. Percept. 3*, 2, 136–151. ACM Press, New York, NY.

Bertin, J. 1983. *The Semiology of Graphics*. University of Wisconsin Press, WA.

BERTRAND, M. AND MULLAINATHAN, S. 2000. Do People Mean What They Say? Implications for Subjective Survey Data. Mimeo, University of Chicago. *http://post.economics.harvard.edu/faculty/mullainathan/papers/peoplemeansay.pdf*. Last accessed 02/08/2006.

BLINN, J. F. 1978. Simulation of wrinkled surfaces. In *Proceedings of the 5th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press, New York. 286–292.

BOLIN, M. R. AND MEYER, G. W. 1995. A frequency based ray tracer. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press, New York. 409–418.

BOLIN, M. R. AND MEYER, G. W. 1999. Visual difference metric for realistic image synthesis. In *SPIE Human Vision and Electronic Imaging IV*, B. E. ROGOWITZ and T. N. Pappas, Eds. SPIE, 106–120.

BRUCE, V., GREEN, P. R., AND GEORGESON, M. A. 1996. *Visual Perception: Physiology, Psychology, and Ecology*. Psychology Press, New York.

CATER, K., CHALMERS, A., AND WARD, G. 2003. Detail to attention: Exploiting visual tasks for selective rendering. In *Proceedings of the 14th Eurographics Workshop on Rendering*, Eurographics Association, Leuven, Belgium. 270–280.

CATMULL, E. 1974. A Subdivision Algorithm for Computer Display of Curved Surfaces. PhD Thesis, University of Utah, Salt Lake City, UT.

CHIU, K. AND SHIRLEY, P. 1994. Rendering, complexity, and perception. In *Photorealistic Rendering Techniques*. Eurographics, G. Sakas, P. Shirley, and S. Muller, Eds. Springer-Verlag, New York, 21–36.

CHUNG, A. J., DELIGIANNI, F., SHAH, P., WELLS, A., AND YANG, G.-Z. 2004. Enhancement of visual realism with BRDF for patient specific bronchoscopy simulation. Lecture Notes in Computer Science (MICCAI). Springer-Verlag, New York.

COOK, R. L. AND TORRANCE, K. E. 1982. A reflectance model for computer graphics. *ACM Trans. Graph. 1*, 1, 7–24.

CORNELL. 2006. Cornell University Program of Computer Graphics—Research Framework for Realistic Image Synthesis. http://www.graphics.cornell.edu/research/globillum/index.html#framework. Last accessed 02/08/2006.

DALY, S. 1993. The visible differences predictor: An algorithm for the assessment of image fidelity. In *Digital Images and Human Vision*, A. B. Watson, Ed. MIT Press, Cambridge, MA. 179–206.

DANA, K. J., GINNEKEN, B. V., NAYAR, S. K., AND KOENDERINK, J. J. 1999. Reflectance and texture of real-world surfaces. *ACM Trans. Graph. 18*, 1, 1–34.

DEMPERE-MARCO, L. 2004. Analysis of visual search for knowledge gathering. PhD Thesis, University of London, London, UK.

DEMPERE-MARCO, L., HU, X. P., MACDONALD, S. L. S., ELLIS, S. M., HANSELL, D. M., AND YANG, G. Z. 2002. The use of visual search for knowledge gathering in image decision support. *IEEE Transactions on Medical Imaging 21*, 7, 741–754.

DUCHOWSKI, A. T. 2002. A breadth-first survey of eye-tracking applications. *Behavior Research Methods Instruments & Computers 34*, 4, 455–470.

DUMONT, O., MASSELUS, V., ZAENEN, P., WAGEMANS, J., AND DUTRÉ, P. 2005. A perceptual quality scale for image-based relighting of glossy surfaces. Department of Computer Science—TR 417, Katholieke Universiteit Leuven, Leuven.

EBERT, D. S., MUSGRAVE, F. K., PEACHEY, D., PERLIN, K., AND WORLEY, S. 1998. *Texturing & Modeling: A Procedural Approach*. AP Professional. San Diego, CA, USA.

ELHELW, M. A., LO, B. P., DARZI, A., AND YANG, G.-Z. 2004. Real-time photo-realistic rendering for surgical simulations with graphics hardware. In *Lecture Notes in Computer Science (MIAR)*, Springer-Verlag, New York. 346–352.

FERNANDO, R. AND KILGARD, M. J. 2003. *The Cg Tutorial*. Addison Wesley, Reading, MA.

FERWERDA, J. A., PATTANAIK, S.N., SHIRLEY, P., AND GREENBERG, D. P. 1996. A model of visual adaptation for realistic image synthesis. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press, New York. 249–258.

FOLEY, J. 2004. The Computer Graphics Grand Challenge: How Much Realism is Enough Realism? http://fp.tm.tue.nl/jfschouten/colloquia/upcoming/foleycv.htm. Last accessed 02/08/2006.

FREEMAN, J., AVONS, S. E., PEARSON, D. E., AND IJSSELSTEIJN, W. A. 1999. Effects of sensory information and prior experience on direct subjective ratings of presence. *Presence-Teleoperators and Virtual Environments 8*, 1, 1–13.

GORAL, C. M., TORRANCE, K. E., GREENBERG, D. P., AND BATTAILE, B. 1984. Modeling the interaction of light between diffuse surfaces. In *Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press, New York. 213–222.

GORTLER, S. J., GRZESZCZUK, R., SZELISKI, R., AND COHEN, M. F. 1996. The lumigraph. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press, New York. 43–54.

GREENBERG, D. P. 1999. A framework for realistic image synthesis. *Commun. ACM 42*, 8, 44–53.

HACISALIHZADE, S. S., STARK, L. W., AND ALLEN, J. S. 1992. Visual-perception and sequences of eye-movement fixations—a stochastic modeling approach. *IEEE Transactions on Systems Man and Cybernetics 22*, 3, 474–481.

HANRAHAN, P. AND KRUEGER, W. 1993. Reflection from layered surfaces due to subsurface scattering. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press, New York. 165–174.

HE, X. D., TORRANCE, K. E., SILLION, F. O. X., AND GREENBERG, D. P. 1991. A comprehensive physical model for light reflection. In *Proceedings of the 18th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press, New York. 175–186.

HECKBERT, P. S. AND GARLAND, M. 1994. Multiresolution modeling for fast rendering. In *Graphics Interface*. Banff, Canada, Canadian Inf. Proc. Soc. Banff, Alberta, Canada. 43–50.

HENDERSON, J. M. AND FERREIRA, F. 2004. Scene perception for psycholinguists. In *The Interface of Language, Vision, and Action: Eyemovements and the Visual World*, J. M. Henderson and F. F erreira, Eds. Psychology Press, New York. 1–58.

HENDERSON, J. M. AND HOLLINGWORTH, A. 1998. Eye movements during scene viewing: An overview. In *Eye guidance in Reading and Scene Perception*, G. Underwood, Ed. Elsevier Science, Oxford.

HENDERSON, J. M. AND HOLLINGWORTH, A. 1999. High-level scene perception. *Annu. Rev. Psychol. 50*, 243–271.

HOWLETT, S., HAMILL, J., AND O'SULLIVAN, C. 2004. An experimental approach to predicting saliency for simplified polygonal models. In *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization*. Los Angeles, California, ACM Press, New York. 57–64.

HU, X. P., DEMPERE-MARCO, L., AND YANG, G. Z. 2003. Hot spot detection based on feature space representation of visual search. *IEEE Trans. Med. Imaging 22*, 9, 1152–1162.

IJSSELSTEIJN, W., DE RIDDER, H., FREEMAN, J., AND AVONS, S. E. 2000. Presence: Concept, determinants and measurement. In *Human Vision and Electronic Imaging V*, 520–529.

ITTI, L., KOCH, C., AND NIEBUR, E. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence 20*, 11, 1254–1259.

KUNDEL, H., NODINE, C., AND CARMODY, D. 1978. Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. *Investigative Radiology 13*, 175–181.

KUNDEL, H. L. 2004. Reader error, object recognition, and visual search. In *Medical Imaging 2004: Image Perception, Observer Performance, and Technology Assessment*. D. P. Chakraborty and M. P. Eckstein, Eds. SPIE. 1–11.

LARSON, G. W., RUSHMEIER, H., AND PIATKO, C. 1997. A visibility matching tone reproduction operator for high dynamic range scenes. *Visualization and Computer Graphics, IEEE Transactions 3*, 4, 291–306.

LAW, B., ATKINS, M. S., LOMAX, A. J., AND WILSON, J. 2003. Eye trackers in virtual laparoscopic training environment. In *Medicine Meets Virtual Reality 11*, 2003, IOS/Ohmsha Press. Amsterdam, The Netherlands. 184–186.

LAWRENCE, J., RUSINKIEWICZ, S., AND RAMAMOORTHI, R. 2004. Efficient BRDF importance sampling using a factored representation *ACM Trans. Graph. 23*, 3, A. B. Watson, Ed. 496–505.

LEVOY, M. AND HANRAHAN, P. 1996L ight field rendering In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press, New York. 31–42

LI, B., MEYER, G. W., AND KLASSEN, R.V. 1998. A comparison of two image quality models. In *Human Vision and Electronic Imaging III*. SPIE. 98–109.

LIU, A., TENDICK, F., CLEARY, K., AND KAUFMANN, C. 2003. A survey of surgical simulation: Applications, technology, and education. *Presence-Teleoperators and Virtual Environments 12*, 6, 599–614.

LONGHURST, P. AND CHALMERS, A. 2004. User validation of image quality assessment algorithms. In *Proceedings of Theory and Practice of Computer Graphics*. 196–202.

LOSCHKY, L. C. AND MCCONKIE, G. W. 2000. User performance with gaze contingent multiresolutional displays. In *Proceedings of the Symposium on Eye Tracking Research & Applications*. ACM Press, New York. 97–103.

LUBIN, J. 1995. A visual discrimination model for imaging system design and evaluation. In *Vision Models for Target Detection and Recognition*, E. Peli, Ed. World Scientific, New Jersey. 245–283.

MACK, A. AND ROCK, I. 1998. *Inattentional Blindness*. MIT Press, Cambridge, MA.

MATUSIK, W., PFISTER, H., BRAND, M., AND MCMILLAN, L. 2003. A data-driven reflectance model. *ACM Transactions on Graphics 22*, 3, 759–769.

MCCOOL, M. D., ANG, J., AND AHMAD, A. 2001. Homomorphic factorization of BRDFs for high-performance rendering. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press, New York. 171–178.

MCMILLAN, L. 1997. An image-based approach to three-dimensional computer graphics. PhD Thesis. University of North Carolina.

MCNAMARA, A. 2001. Visual perception in realistic image synthesis. *Computer Graphics Forum 20*, 4, 211–224.

MCNAMARA, A., CHALMERS, A., TROSCIANKO, T., AND GILCHRIST, I. 2000. Comparing real and synthetic scenes using human judgements of lightness. In *Proceedings of the Eurographics Workshop on Rendering Techniques*. Springer-Verlag, New York. 207–218.

MITCHELL, D. P. 1987. Generating antialiased images at low sampling densities. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press, New York. 65–72.

MURPHY, H. AND DUCHOWSKI, A. T. 2001. Gaze-contingent level of detail rendering. In *EuroGraphics [short presentations]*. EuroGraphics Association. Manchester, UK.

MYSZKOWSKI, K. 1998. The visible differences predictor: Applications to global illuminations problems. In *Eurographics Workshop on Rendering Techniques*, G. Drettakis and N. Max, Eds. Springer-Verlag, New York. 223–236.

NICOLAOU, M., JAMES, A., DARZI, A., AND YANG, G.-Z. 2004. A study of saccade transition for attention segregation and task strategy in laparoscopic surgery. In *MICCAI*. 97–104.

NODINE, C. F. AND KUNDEL, H. L. 1987. The cognitive side of visual search in radiology. In *Eye Movements: From Psychology to Cognition*, O. R. JK and L.-S. A, Eds. Elsevier Science, Amsterdam. 572–582.

OHSHIMA, T., YAMAMOTO, H., AND TAMURA, H. 1996. Gaze-directed adaptive rendering for interacting with virtual space. In *Proceedings of the Virtual Reality Annual International Symposium (VRAIS 96)*. IEEE Computer Society, Los Alamitos, CA. 103–110.

OST, D., DE ROSIERS, A. , BRITT, E. J. , FEIN, A., LESSER, M., AND METHA, A. 2001. Assessment of a bronchoscopy simulator. *American Journal of Respiratory and Critical Care Medicine 164*, 12, 2248–2255.

O'SULLIVAN, C., DINGLIANA, J., AND HOWLETT, S. 2003. Eye-movements and interactive graphics. In *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movement Research*, R. Radach, J. Hyona, and H. Deubel, Eds., Elsevier Science, Oxford. 555–571.

PERLIN, K. 1985. An image synthesizer. In *Proceedings of the 12th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press, New York. 287–296.

POMPLUN, M., REINGOLD, E. M., AND SHEN, J. Y. 2001. Investigating the visual span in comparative search: The effects of task difficulty and divided attention. *Cognition 81*, 2, B57–B67.

SRADEMACHER, P., LENGYEL, J., CUTRELL, E., AND WHITTED, T. 2001. Measuring the perception of visual realism in images. In *Proceedings of the 12th Eurographics Workshop on Rendering Techniques*. Springer-Verlag, New York. 235–248.

RAYNER, K. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin 124*, 3, 372–422.

RENSINK, R. A. 2000. Seeing, sensing, and scrutinizing. *Vision Research 40*, 10–12, 1469–1487.

RUSHMEIER, H., WARD, G. J., PIATKO, C., SANDERS, P., AND RUST, B. 1995. Comparing real and synthetic images: Some ideas about metrics. In *Eurographics Workshop on Rendering Techniques*, P. M. Hanrahan and W. Purgathofer, Eds. Springer-Verlag, New York. 82–91.

SIDAK, Z., SEN, P., HAJEK, J. 1999. *Theory of Rank Tests (Probability and Mathematical Statistics)*. Academic Press, New York.

SILLION, F. AND PUECH, C. 1989. A general two-pass method integrating specular and diffuse reflection In *Proceedings of the 16th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press, New York. 335–344.

SPSS. 2006. http://www.spss.com. Last accessed 02/08/2006.

STOKES, W. A., FERWERDA, J. A., WALTER, B., AND GREENBERG, D. P. 2004. Perceptual illumination components: A new approach to efficient, high quality global illumination rendering. *ACM Trans. Graph. 23*, 3, 742–749.

SUNDSTEDT, V., DEBATTISTA, K., LONGHURST, P., CHALMERS, A., AND TROSCIANKO, T. 2005. Visual attention for efficient high-fidelity graphics. In *Spring Conference on Computer Graphics (SCCG 2005)*. 162–168.

TOBII. 2006. Tobii Technology User Manual. http://www.tobii.se. Last accessed 02/08/2006.

TUMBLIN, J. AND RUSHMEIER, H. 1993. Tone reproduction for realistic images. *Computer Graphics and Applications. IEEE 13*, 6, 42–48.

VELICHKOVSKY, B. M., DORNHOEFER, S. M., PANNASCH, S., AND UNEMA, P. J. A. 2000. Visual fixations and level of attentional processing In *Proceedings of the Symposium on Eye Tracking Research & Applications*. ACM Press, New York. 79–85

WALLACE, J. R., COHEN, M. F., AND GREENBERG, D. P. 1987. A two-pass solution to the rendering equation: A synthesis of ray tracing and radiosity methods. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press, New York. 311–320.

WARD, G. J. 1992. Measuring and modeling anisotropic reflection. In *Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press, New York. 265–272

WARD, G. J. 1994. A contrast-based scalefactor for luminance display In *Graphics Gems IV*. Academic Press Professional. 415–421.

WATT, A. AND WATT, M. 1992. *Advanced Animation and Rendering Techniques: Theory and Practice*. ACM Press, New York.

WHITTED, T. 1980. An improved illumination model for shaded display. *Commun. ACM 23*, 6, 343–349.

Yang, G. Z., Dempere-Marco, L., Hu, X. P., and Rowe, A. 2002. Visual search: Psychophysical models and practical applications. *Image and Vision Computing 20*, 291–305.

Yarbus, A. L. 1967. Eye movements during perception of complex objects. In *Eye Movements and Vision*, L. A. Riggs, Ed. Plenum Press, New York. 171–196.

Yee, H., Pattanaik, S., and Greenberg, D. P. 2001. Sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Trans. Graph. 20*, 1, 39–65.

Received June 2005; revised January and August 2006; accepted January 2007