A METHOD AND SYSTEM FOR INTERACTION WITH MICRO-OBJECTS

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to the fields of microscopy and digital interaction and display.

2. Background of the Art

Seeing small things enlarged, and working on them, began with the lens. A large lens automatically presents paired distinct views (a stereo image) to the two eyes of the operator, and a change in view (parallax) when the operator's head moves. These two depth cues are among the most powerful for the human vision system, and enable a strongly three-dimensional sense of the object looked at. This is an important help to the operator in manipulating objects, since any tool must be aimed correctly in depth, as well move correctly across the image in various directions, for proper effect.

In about [*date*] the microscope (combining multiple lenses and other elements) brought greater enlarging power, at the cost of reduced depth cues. Since the operator must place an eye exactly at the eyepiece, the head motion needed for parallax is made impossible. The standard single-eyepiece microscope also bars stereo, by providing only one view. The operator is left with weaker depth cues such as occlusion of a more distant object by a near one, without the quantitative precision need for aiming a tool.

The invention of the binocular microscope in [*date*] merely added comfort, since both eyes saw the same view: there was no need to cover or shut one eye, or ignore what it saw. The brain could integrate a sharper combine view, but it could not reconstruct depth, and guiding microtools remained hard. The step to the stereo microscope in [*date*] restored binocular depth perception, but parallax remained unavailable. Since stereo depth perception is strong only in about half the population, This delivered good depth only to approximately half of the population. Surgeons for whom stereo is a strong cue were empowered as microsurgeons, but those perceiving depth primarily in other ways could not use the stereo microscope easily: not only could they not make microsurgery a specialty, but where some part of a mostly-larger-scale procedure could benefit from enlargement the available tool was not useful to them. Similar observations apply to other tasks of micromanipulation, such as micro-assembly of electronic parts.

However, the surgical microscope did continue one key merit of the simple lens: hand-eye co-location and co-orientation. In a direct view, the visually perceived position of a hand or tool coincides almost exactly with the neuromuscular sense of where it is, and

in what direction its parts are pointed, through the 'proprioception' of the operator's own joint positions. These are key parts of the awareness needed in guiding a tool. Enlargement necessarily sacrifices the locational match to some extent, since the ends of a one-millimetre movement that looks like a centimetre cannot both match feel and view, but the scene can still be centred in a common location. More importantly, every rotation has an axis direction and an angle turned through, and these can match even between touch and a magnified view. This correspondence is exactly preserved (Drawing 1) in some surgical microscopes 100, with an objective lens 101 and an eyepiece 102 through which the operator's eye or eyes 105 (shown in profile) can see vertically down from an eyepieces or eyepieces 101 facing exactly upward, seeing an enlarged view of the work area 106. However, most work is more comfortable with a tilt. The system of lenses and mirrors inside an indirect microscope 120 bends to deliver the view of the work area 126 from a vertical object lens assembly 121 to an eyepieces or eyepieces 122 at a lower angle, so that the eye or eyes 125 of a sitting operator can look in. The orientations match imperfectly, and the visual location 129 (apparent distance depends on stereo adjustments) of the operand is moved substantially from its physical site 126. Operators learn to overcome these obstacles, but they are non-trivial.

In contrast [*details of our first patent application*], hereby incorporated by reference and hereinafter cited as "PLR", introduced the system 150, where the work site and its contents appear to the user in substantially the same location and orientation as they appear to the user's motor cortex. (As with the surgical microscope, a variant exists for two collaborating operators.) We will refer to such a system, including the extensions described below, as a **flatscope**. A display appears on a thin flat screen 151, which becomes a stereo view by shutter glasses 152 synchronised with the alternating left and right eye views on the screen 151, or by a means of directing different view in different directions. The subject 155 of the operator's actions, here illustrated as a pair of artery ends to be joined, appears enlarged 156 in the display 151, and any tool 160 held by the operator in the work area is similarly enlarged 161. The near-match of apparent to real location, and the precise match of apparent to real orientation, eases the neuromotor tasks of the operator. The replacement of an optical eyepiece by a larger display is not unique to the system 150, but the close match between hand and eye in perceived location is missing in such systems as [*James insert some very-up-to-date examples*], which place the screen away from the line between the eye and the physical operand, and typically rotate it by as much as a right angle. We exclude systems with such major deviation from hand/eye co-location from our use of the term 'flatscope'.

The purpose of the present application is to disclose some substantial additions to the system 150, made practical and economic by developments since the original invention. The long-barreled cameras whose geometry was there described as compensated by mirrors can now be replaced by cameras flat enough to be placed directly under the display 151, facing the work area 155. An example of such a thin system which enables "a 3D depth map without the need for active illumination or calibration" is described by K Fife, A El Gamal and H-S P Wong, *A 3MPixel Multi-Aperture Image Sensor with 0.7um Pixels in 0.11um CMOS*, IEEE ISSCC Digest of Technical Papers, pp. 48–49, February 2008. The small differences between the light captured by the different apertures in that device described allow computation analogous to the use of greater differences in the views by physically separate eyes or cameras, after which each pixel of the processed image has a depth associated with it. It is not of the essence of the present disclosure that the specific method there

described be used, but it illustrates that images with depth data can be captured from a single device, with a size and fabrication technology which make consumer products probable. We may thus assume this capability in image sensing devices to be incorporated in a flatscope. Alternatively, depth may be determined by a variety of methods well known to those skilled in the art, such as computed stereo from two or more standard images, the use of structured light, or active devices depending on the timing of a reflected signal, such as ranging by laser or ultrasound devices. As we shall show, taken together with the increased speed and reduced cost of computation, such data allow the addition of the parallax depth cue to a flatscope, enlarging the range of operators for whom the system is natural and usable with minimal training. Such advantages over eyepiece microscopes as the reduction of strain, since the operator is not compelled to keep head and shoulders in a fixed position throughout any procedure, become even stronger in the presence of parallax, since the apparent position of the operand is more stable than with stereo alone.

Similarly, the sensing of images in more than three overlapping bands corresponding roughly to ranges which stimulate red, green and blue receptors in the human retina — whether by extending into the infrared range, as in USPTO A20060097176, Szu, *Infrared multi-spectral camera and process of using infrared multi-spectral camera,* October 2008, for example, or USPTO A20080260225, Szu, *Infrared Multi-Spectral Camera and Process of Using Infrared Multi-Spectral Camera*, October 2008, or into the ultraviolet, or subdividing the visible spectrum more finely — has become practical at an economic and technical level appropriate for wide use in medicine and other fields. The present invention does not concern the hardware or software needed for multi-spectral sensing or analysis as such, but builds upon the practical availability of such multi-spectral image data as it does on the availability of depth data.

BRIEF DESCRIPTION OF THE INVENTION

TBD

BRIEF DESCRIPTION OF THE FIGURES

1. Vertical and indirect microscope designs, and a flat display for magnified work.
2. Profile views of a single user flatscope, and a two-user flatscope.
3. Disposition of multiple cameras and lights on the underside of a flatscope.
4. Eye position dependent display of an enlarged object in a flatscope.
5. A support frame adding some flatscope functions to a normally hand-held device.

DETAILED DESCRIPTION OF THE INVENTION

Image acquisition and display to hand and eye

As shown in Drawing 1, between the operator's physical work area and the operator's eyes is placed a display screen, on which a magnified view of the work area appears to the operator in substantially the same location and orientation as they appear to the user's motor cortex. We refer to this as 'hand/eye co-location', though the operator often inserts hand-held tools into the work area, seeing the hand positions via tool positions, rather than directly. The material in the work area on which the operator acts as the 'operand'. Typically the magnification is up to fifteen times, not because of

technical obstacles to higher magnification but because hand tremor makes a closer view unhelpful. With a means to reduce tremor, which is not the subject of the present disclosure, the usable magnification factor increases: in certain cases a factor of 1 or less can be useful. With indirect means of controlling tremor, such as the use of remotely controlled robot arms at a distant location, the hand/eye co-location is specifically between where the work site is seen to be and where it is felt by the hands to be, without regard to where it physically is. In a preferred implementation, the display screen is substantially orthogonal to the line of view, but appropriate construction of the view permits any angle at which the screen wholly covers the eyes' view of the felt work area and can be clearly seen.

Two operators may collaborate on a shared magnified operand: this is particularly common among surgeons, where tasks such as suturing are easier with more than two hands. As in Drawing 2, a single screen 201 may be intermediate between the physical work area 210 and the eyes 205 of a plurality of operators (in our preferred embodiment, two operators), or the display 211 may place two flat screens between the work area 210 and the eyes 215 of two operators. Analogous placements for three or four operators will be evident to persons skilled in the art.

Hand/eye co-location involves not only the visual sense of the direction in which the operand lies, but the distance from the eyes of each object within the work area, referred to as 'depth'. Some sensory cues of depth are relative, such as which object occludes another and is thus closer to the eyes. Accurate manipulation requires awareness of the absolute position in all directions. The main visual cues for absolute depth are stereo, reconstruction of distance from the differences between the views of the two eyes, and parallax, reconstruction of distance from the changes in view as the head (and thus each eye) moves. The disclosure PLR relies on an arrangement of a pair of cameras (or one camera with a split view) and optics so that the views captured geometrically correspond to the views that would appear to eyes in the expected location. This approach is viable, but limited: in particular, unless the optics move as the eyes move (which is not impossible, but creates substantial difficulty, and requires mechanical motion of parts), the approach presents two fixed views, making parallax impossible. We here disclose a more 'computational camera' approach by which most of the geometry is handled in software.

Under the display surface 151 (presented less schematically as 300 in Drawing 3) we locate one or more camera units 301, and one or more lights 302. Drawing 3 emphasises that there may be many more than one of each, but any arrangement that provides shadow-free lighting and depth data is within the spirit of the present of the present invention. As mentioned in the Background to the Invention, even a single multi-aperture chip can now provide depth data, but wafer-level optics has reduced not only the size but the price of visual sensors to the point where minor advantages can justify their use. If two cameras are used, their images may be computationally compared so as to derive depth data (alternatively to, or enhancing the precision of, single-camera depth reporting), or simply arranged as a pair of stereo views reaching the operator's eyes. An operator with a good stereo depth perception then acquires a clear sense of the depth of different parts of the displayed region, appropriately magnified if the angular separation of the camera positions is appropriate: but this is not our preferred embodiment, since explicit depth data are required for the support described below for the parallax depth cue, and for other enhancements of the system.

We also include, in our preferred embodiment, one or more video cameras whose field of view includes the expected range of positions of the eyes of the operator or operators. The output of this or these cameras goes to a digital processor, which as one of its activities identifies the positions of the operator's eyes within the images. (Various means of doing this are well known to those skilled in the art.) If there are a plurality of cameras, triangulation then establishes the distance of the eyes as well as their direction from the cameras, and thus their full three-dimensional location in space relative to the flatscope. If there is only one camera, we specify it as a camera that can capture depth, so that again the three-dimensional location of the eyes becomes available. If more than two cameras are installed, or a plurality of cameras with depth reporting, the redundancy may be used to improve the accuracy of the position data for the eyes.

These eye-position data may be used for control of the system by the user in various ways, as described below, but an important feature is that this combination of depth data for the image of material below the flatscope, and location for the eyes above it, permits the return of the parallax depth cue, which has been absent from any enlargement system more complex and powerful than a simple lens. Using pixel coordinates (i,j) for the display, with the current digital magnification, we have colour and depth data at spatial co-ordinates $(i,j,d(i,j))$, showing a view of the visible surface from a specified point for an eye or virtual camera. Drawing 4 shows a flatscope 400 with a single depth-reporting camera 405 centred on its underside: other locations, or multiple cameras as described above, may equally be used. Below it is an operand 409, which the operator is to see as a larger object 419 by means of appropriate imaging on the upper surface of the flatscope 400. With the operator's eye (or eyes, seen in profile) at the position 413 vertically centred above the flatscope 400, the view 420 from the camera 405 would be approximately correct for the intended apparent size and location of the virtual object 419. However, for an eye or eyes in the location 411, three points equally horizontally spaced on the virtual object 419 should be rendered at the points 415 shown, on the direct lines between the points on the virtual object 419 (the enlarged operand) and the eye location 411. This correctly locates their images, as seen from the position 411. Due to the unequal separations between such points, the optimal image to be rendered on the flatscope display is as illustrated in the view 420. This view may straightforwardly be produced by computational methods well known to those skilled in the art, provided the depth information is available.

Alternatively, a tracking device fitted to and moving with the operator's head may provide such eye-position data: if the user is wearing shutter glasses 152, or other such support for stereo display, the tracking device may be attached to or embedded in the structure of the glasses.

It is important that sufficient computing resources be available to deliver the image depth and eye location data accurately in a short time, preferably substantially shorter than the refresh interval (typically of the order of 20 milliseconds) of the display, so that the refreshed image is redrawn using appropriate data. If the image shown changes correctly with head motion (as known to the operator by sensors in the inner ear), parallax provides a strong sense of a stably located operand 419 at a well defined position, even in the absence of the stereo depth cue (which for some operators is always absent or weak). Inaccurate or out-of-date eye locations make the object appear to jitter in space, with results less useful than a simple display of the untransformed view from a camera at the location 405.

Tracking data should be available to the processes described below with a latency below 20 milliseconds, where latency is defined as the time between the device being physically in a location and the time that location is delivered to the rest of the system, not (as often defined by tracking device manufacturers) the time to respond to a query with the most recent position computed, at an unspecified physical time. In both cases, prediction based on recent data may be used to improve the estimate of eye position at the time of the next refresh of the screen, and this estimate used in preference to the recent positions directly measured or computed. The appropriate view 430 may be rendered fast as a 'terrain' by methods well known to those skilled in the art, without the need to construct a mesh of triangles on the fly for standard general surface rendering.

An advantage here of using multiple cameras is that fewer points hidden from a camera location like 405, but visible from an oblique eye position like 411, must be rendered by interpolation (computational guessing): every point that should be 'visible through' the flatscope 400 is visible to a camera at some location under it. There is some added complexity to the depth data structure, since a sideways view can include points overhung by others, and consequently to the rendering algorithms, but many ways of dealing with these will be evident to those skilled in the art, within the spirit of the present invention. We note that all the processes discussed above, whether in determining depth and eye data or in rendering, are extremely compatible with parallel computing methods that fit well in the pipeline of graphics processing units (GPUs) with hundreds of parallel cores, that can now be programmed directly using programming language extensions such as CUDA (Compute Universal Device Architecture).

In an alternative embodiment, where the flatscope uses multiple cameras and the direction of view from the current position of the eye or eyes is intermediate between the directions in which two or more cameras view the work area, interpolation between the views reported by these cameras gives an image in the appropriate direction, by planar image manipulation methods. However, this does not take into account the changing distance of an eye from the display, and thus cannot present as stable an apparent operand 419 as the use described above of explicit depth data.

Imaging modalities

Light of the frequencies used by cameras (the visible spectrum, plus the near infra-red and ultra-violet) is only one of the modalities by which a depth-including image may be generated for display on the flatscope. Any real-time image acquisition system may be used, including but not limited to the following, underlined in the paragraphs below:

An ultra-sound (US) signal is emitted by one or more emitters attached to or embedded in the flatscope, or alternatively placed near against the part of an object that is under the display, such as a human or an animal, optionally with gel or other means to reduce impedance between the emitter and the object. Sensors for echoed sound are placed similarly, optionally as part of the same device or devices. In the case of sensors and emitters moved separately by the operator or moving with the operand, the emitters and sensors are tracked in real time (by analysing the images from the camera unit or units 301, or by other means known to those skilled in the art), so that each element of the US-derived 3-dimensional image has a known location relative to the flatscope. The flatscope is thus enabled to display the US-derived image in a stable view, with hand-eye colocation, using any convenient magnification factor. In a preferred embodiment the US-derived image is fused with the visual image, so that if the operand is a patient the effect is of a view through translucent skin. This embodiment is of particular utility for minimal access procedures, where the operator inserts small tools

through holes in the skin, rather than create a large opening for a direct view: examples include but are not limited to biopsy, laparoscopy and arthroscopy. The operator in such a procedure must with many current technologies look at a display in a location far from the operand, and variably rotated relative to it: an ultra-sound screen is often placed to one side of the operator, with a view whose orientation varies with the hand holding the device; a laparoscopy display in front of the surgeon shows the view from a movable camera inside the abdomen, several feet below the screen, and including indirectly controlled tools; and so on for other cases. The resulting difficulty of hand-eye coordination increases error rates, strain on the operator, time to complete a procedure, and cost. In many such applications the flatscope embodiment just described can mitigate this difficulty, and in some cases replace the inserted camera as an imaging means, reducing the invasiveness of a procedure.

A flatscope constructed of MRI-compatible material (without electromagnetic effects that interfere with magnetic resonance imaging, and not harmed by the large magnetic fields used), used in an open MRI system, allows an operator hand-eye colocated manipulation of an operand imaged using magnetic resonance. As with US, in our preferred embodiment this image is fused with a camera view.

Continuous or frequent exposure to X-rays, and thus to computerized tomography (CT), is unwise, but interventional radiography uses X-ray snapshots to guide catheters and other minimal access tools. Often an injected contrast agent makes arteries or other structures visible. Views from two directions can suffice (as in stereo vision) for 3-dimensional location, and hence a 'road map' in three dimensions rather than in a single projected view. If this is fused with a visible display in a flatscope, the operator's manipulation becomes easier.

If the patient or other operand moves or is moved during a procedure, the displayed position of the arteries becomes outdated. An embodiment of the present invention uses the geometric data from the cameras to construct a model of the bending and motion of the operand, and uses this to update the displayed X-ray data. More generally, such geometric update may be applied either to recent data acquired in situ, from X-ray, MRI, PET or any other scanning method, and adjusted to the geometry revealed by the visual component of the flatscope.

The image displayed by the flatscope may also be derived from simulation, where a good enough interactive model of the operand is available: see 'training system' below.

The image displayed by the flatscope may also be derived from a remote site, where a camera scheme compatible with that of the flatscope exists. In our preferred embodiment, this is another flatscope, but the flatscope operating system can accept image and annotation information in widely used formats such as DICOM, and thus support information from any remote site that implements industry standards. Similarly, the images and annotation collected by a flatscope and constructed by an operator are exportable in real time to any standards-compliant device. If remotely controlled tools are available at a site which is thus receiving an image of the operand and transmitting responses, the remote operator can then act directly upon the operand: if they are not, the remote operator's local tools can be used for pointing, annotation, *etc.*, creating signals that add transiently or otherwise to the view seen by the operator on the spot. This may be used for mentoring and consultation.

The image displayed by the flatscope may also be derived from a recording, of an earlier procedure or an earlier part of the current procedure, played back at the original speed, or faster, or slower. This may be used to watch a procedure or a step in more

detail, to analyse the errors of oneself or another operator, or for emulatory training: the current operator's tools are added to the view of a procedure performed earlier by an expert, and (optionally starting with a slowed version) the current operator attempts to reproduce the motions of the expert. The system measures and reports on the success of this attempt.

User interface (UI)

As in the example of annotation below, for a flatscope, user input is often better with the hands below the display, rather than on top (as with a touch-screen) or away to one side (as with a mouse). One reason to avoid screen-touching during a procedure is that a surgical or other physical operand is a rich source of material which could reduce visibility. 'Keyboard plaque' is rarely bad enough to hide the letters on the keys — and those who use a keyboard that much can usually touch-type — but any visible smear on the screen is a serious problem. In our preferred embodiment, therefore, we reserve screen touching to initial tasks of setup and calibration. [*insert something on this later*]

Unlike the planar input of any screen-oriented interaction, whether an object or cursor is moved co-locationally by tracking touch, or indirectly by a mouse or other device, the flatscope has co-locational tools as part of its core design. The operator uses these delicately to act on a physical operand, and it is natural to extend them to control of the flatscope. The operator's hands need not change to a different control object, nor to a different space such as a mouse-pad or the front of a screen. The tip of a scalpel, forceps, engraving tool or other implement held by the operator moves with the full six degrees of freedom normal for a rigid object: translation back and forth, side to side, up and down, and rotation about any axis, including the forward, lateral and vertical axes. The human operator is accustomed to manipulating these six degrees of freedom simultaneously (for instance, bringing a cup along a curved path to the mouth, while rotating it to be ready for drinking), unlike combining a scroll-wheel with the 2D input of a mouse, or managing six-degree-of-freedom motion via two degrees of freedom in the separate positions of three fingertips.

To use these tool tips as input devices, it is necessary for the flatscope system to 'know' (contain explicit data specifying) their positions. Various tracking devices well known to those skilled in the art may be attached to such tools, including but not limited to a robot arm, a receiver of radio-frequency signals or ultrasound from a fixed emitter, light-emitting diodes or other point sources of light to be detected by a camera, and so forth, and any of these may be used within the spirit of our invention. In our preferred embodiment, however, any pointed rigid object that enters the field of view of the camera or cameras in Drawing 3 may serve for input, without special attachments. Computer vision does not yet match the human system in general object identification in arbitrary scenes, but finding rigid objects of a known shape (here, one shape in a small tool database of the shapes used in a particular field of flatscope use), in front of the scene rather than partly occluded, and with controlled lighting, is a task for which many solutions are known to those skilled in the art: an adequate overview appears at http://en.wikipedia.org/wiki/Computer_Vision. We therefore assume that this positional information is available, in real time, to the computing system of the flatscope. The system can use this to base actions upon the computed answer to "where is the tip?", "where on the operand is the tip pointing?", "what point on the operand is closest to the tip?", "what points of the operand are within the current active radius of the tip?" "what points are in the simulated airbrush effect region of the tip?", "what gesture is the tip making?" (where a gesture is a learnable motion, such as moving the tip in a small

circle), and many other decision actions that will be evident to those skilled in the art, within the spirit of the present invention.

A tool may be in 'physical only' mode, where its effects are only those resulting from physical interaction with the operand, such as cutting, cauterizing, grasping, and so on. Alternatively, it may be in selection mode, painting mode, false-coloring mode, or other modes described below or added within the spirit of the present invention. The operator causes it to enter or leave a mode by voice interaction (speaking a standard word or phrase, detected by a microphone connected to the flatscope), placing the tip near or in a 'menu item' region added to the display by the flatscope system, making a recognized gesture with the tip (for example, a small clockwise circle) when the tip is visible at a distance from the operand, pressing a foot button, or other means that will be evident to those skilled in the art, within the spirit of the present invention. One or more buttons may be added to the hand-proximal part of a tool, with similar effect, but our preferred embodiment uses only the operator's tool 'as is', without redesign or additions, so that addition of a new tool requires only the inclusion of its shape in the tool database: this may be imported from its manufacturing computer-aided design (CAD) specification, or from a three-dimensional scan of an example of the tool, by means well known to those skilled in the art. The effect of any of these mode changing actions may depend on context, just as the effect of an On/Off button depends on whether a device is currently active.

A subclass of tools including needle holders, forceps, tweezers, scissors and other devices, biopsy needles with active capture at the tip, and many others, have visually recognizable states of being 'open' or 'closed' at the distal end, with this state being under the control of the operator (typically by relative motion within the operator's grip, such as separating or joining the handle loops of scissors). Recognition of these states permits a further class of control mechanisms, including a three-dimensional version of the familiar computer-mouse 'drag and drop': the operator closes the tip at or near a particular element or location of the scene, and moves or extends that element or a new element to the three-dimensional location where the tip is opened. This may be used for simple transfer, for example of a tag indicating the next point for cauterization or anastomosis, or for 'rubber banding' which creates a line starting point where the tip closes, and an end point for that line where the tip next opens. For curved visible paths, a 'grab by closing, drag, release by opening' mode allows the operator to change the shape of the path. Unlike the typical curve-dragging familiar to computer mouse users, which moves an interpolation or control point (x,y) to a new point (x',y') without direct control of the direction of the path, the six-degree-of-freedom input natural to a tool in the flatscope allows the operator not only to move the tip from a point (x,y,z) to a new point (x',y',z') and have a path control move with it, but to turn the tip while moving, so that the direction of the path through the point turns similarly. Fewer motions are thus needed to specify a complicated path, such as instructions by a mentor to a trainee or robot as to the safe way another tool should move.

Eye-controlled zoom

Beside the hand and voice controls described above, our preferred embodiment includes control by position of the operator's head and eyes. We have already discussed the use of these data to display images with appropriate parallax, perspective and stereo depth cues, for a display object 419 with a given virtual position. It follows that if the operator's head approaches the display, the object 419 will take up more of the operator's field of view, just as with bringing one's eyes closer to a book: but if only this mechanism applies, the book and object 419 do not appear to be larger.

The human vision system adjusts to see them as close, but with the same size. We are here discussing a change in perceived size.

With zoom mode turned on, by voice or a foot pedal or a menu choice or other such means, the eyes' approach to the flatscope results in a spatial enlargement of the display object 419, so that it appears actually larger, for examination in greater detail: moving away reverses this effect. The factor by which movement translates into an enlargement ratio λ is set by default, for example a=7 centimeters of approach producing a doubling of apparent size ($\lambda=2^{a/7}$), but is modifiable by operator preferences. The default itself is reached by study of a representative group of operators (globally or within an institution), or by a consensus procedure.

The simplest implementation of this zoom process replaces each display location (x,y,z) of a point in the operand or on a tool by $(\lambda x,\lambda y,\lambda z)$. This always transforms (0,0,0) to the point (0,0,0): this point is fixed (like the centre of a Google Maps display during zoom), and other points move radially toward or away from it. In our preferred embodiment the fixed point (X,Y,Z) is under user control and the formula becomes

$$(x,y,z) \rightarrow (X,Y,Z) + (\lambda(x\text{-}X), \lambda(y\text{-}Y), \lambda(z\text{-}Z)).$$

The operator may choose the point (X,Y,Z) by holding a tool tip there, or by fixing the gaze of the eyes on it. Means well known to those skilled in the art detect the direction in which the eyes are looking, as well as the point from which they look, and compute either the point where the gazes of the two objects meet, or the point where the gaze direction meets a displayed element of the operand 419 or a tool visible within it. This point becomes the (X,Y,Z) used in the formula above. Since this point does not move in any direction, an object centred there does not appear to move closer or further away, but does appear to grow or shrink.

The factor λ may be a single number, controlled as in the example above by a formula such as $\lambda=2^{a/7}$ responding to the change a in operator distance, but the same for each (x,y,z) that is zoomed. Alternatively, λ may monotonically diminish with distance, by a formula such as but not limited to $\lambda=2^{a/7}/(1 + c((x\text{-}X)^2 + y\text{-}Y)^2 + (z\text{-}Z)^2))$, where c fixes the rate of diminution. This gives a localized zoom, by which the region around (X,Y,Z) swells, an intermediate zone is radially compressed, and remote points barely move. At the expense of some distortion, the zoomed region is still seen in the same lager context, and navigation is easier.

The use of a single or variable zoom factor λ in display is well known to those skilled in the art, but λ is usually controlled by a slider or other draggable element of the display or its surroundings. Our use of head motion to modify λ appears to be novel either in the context of a three-dimensional display such as the flatscope or other system for examining images with depth, or (omitting z and Z terms from the formulas above) in the context of a system for working with planar images.

Annotation, mensuration and analysis

If the operator signals (by a voice command, gesture, menu choice, or other means known to those skilled in the art) that the tip of the currently held tool is to become an annotator, it may become a brush, pencil, curve creation tool, or other marking widget as familiar in applications such as PhotoShop™. However, whereas in such applications such widgets create a set of colour values associated with pixels in a planar image, in the flatscope the data created include depth values, controlled by the positions of the tip as the creation take place, and become associated with the flatscope's internal representation of the operand 419. Beside colour markings, the values may include tags such as labels for particular points or substructures of the

operand, and classifications such as 'necrotic' or 'not to be cut', which may be displayed by modifying the colour in which a set of pixels is displayed, but are not to be identified with colour values. (Among other consequences of this difference, the system can permit the operator to choose to show necrotic points as darkened, lightened, blue, *etc.*, as convenient.) In a preferred embodiment the marking mode is set by the user, either attaching values to particular points in the work area, or to points on the surface of the operand: in the latter case the marking may be of the points behind the tip relative to the operator's current viewpoint, of the points nearest the tip, of the points reached by a virtual airbrush, or of points selected by other criteria within the spirit of the current invention. In the former case the operator creates free-form structures in space, which are displayed as occluding the operand 419, opaquely or translucently, or as occluded by it, according to the needs of a particular application. In a preferred embodiment, all such display is adjusted to the point of view of the operator, as shown in Drawing 4.

As already noted under Imaging Modalities, the system may create a deformable model of the operand, which adjusts its position to match the current image data: if the operator cuts or otherwise modifies the operand, the model must change topology as well as deform. In our preferred embodiment the annotations of points on or in the operand move with the locations of points in such a model, so as to appear as though painted or drawn on the operand itself, as seen through the flatscope.

Many important annotations are semi-automatic, in that the operator sets the purpose and scope, but the assignment of annotation range to particular parts of the operand is computational. For example, the user may select a point (see above, UI, for means of selection) and by voice, menu, gesture or otherwise request the system to identify the structure (tube, artery, lesion, *etc.*) that it is part of. The system highlights the identified structure. If this is larger or smaller than the desired object, the operator may request "larger" or "smaller" by voice, identify the kind of structure sought (for example, by saying "tube"), or use detailed editing tools to change the displayed boundary and hence the selected structure. The operator can then request the computation of the structure's area, length, diameter, thickness, rugosity (rough or smooth, *etc.*), degree of inflammation, estimated blood flow, and so on, according to the analysis capabilities in the available embodiment of the present invention. Among the desired capabilities is the quantification of oedema: where a swelling visibly interrupts the apparent shape of a tissue, a surface fitted by minimizing a pseudo-energy constructed by combining measures of bending and of distance from the unaffected region gives a referent against which the volume of swelling can be quantified, as the region between the interpolated and actual surfaces.

As noted in the Background to the Invention, it is increasingly possible and economic to acquire images with considerably more spectral information than the standard set of three colour bands, red green and blue. Just as visible colour allows discrimination beyond what is possible in a grey scale image, increased spectral range and refinement enable machine recognition of material properties not apparent to the naked eye, such as temperature or the spectral characteristics of a particular type of infection with a large yield of (for example) a molecule highly reflective in specific sub-bands of 'red' and 'blue'. Other phenomena such as a bruise can give the same 'mauve' sense to the human eye, by stimulating 'red' and 'blue' to the same total extent, but be easily distinguished by their reflectivity in sub-bands. An application-specific flatscope such as one intended for microsurgery includes in its programming the signatures of a wide range of such relevant signatures, so that the operator can call on the system to identify

one of them and mark (in human-visible pseudo-colour) the region in which it is present. In our preferred embodiment a background routine automatically checks for the spectral characteristics of a default check list of properties that are clinically or otherwise important in the particular application, and alerts the operator to their presence. With cameras sensitive in the infra-red, the system can also deduce temperature, and derive measures of inflammation from this and other spectral contrasts with surrounding tissue. Upon operator request, these measures may be overlaid either on the entire surface by a graphic device such as false colour or contours, or only upon the region where abnormal values appear. Other imaging modalities such as ultrasound make it possible to derive physical quantities such as surface elasticity, which may also be added as visible annotations.

A frequent action by microsurgeons is to press upon an artery, observe the time course of colour after the blockage of blood flow (red colour is lost: blood that does not drain turns blue, indicating strength of drainage), then release it and observe the time course of return to the status quo. Similar actions occur in any field where the operator acts on a distributed object with its own internal dynamic. In most cases, like the blood flow example, this observation is currently unquantified, but the flatscope may be instructed to observe and record the events, attaching the data to a tag for the intervention point. Automatic multi-spectral analysis maps the region that changes, and estimates for each part of it the degree to which it is influenced. (In the blood supply case, this marks how completely each part of the tissue depends on the specific artery branch that is blocked. Most points in tissue have at least some redundancy in supply.) Material or biomedical properties derived from these measures can also be added as visible annotations.

Wherever from one or more of the above measures the system can detect or quantify the extent of quantities significant to the operator, including but not limited to skin elasticity, necrosis, vascularization, infection, perfusion, oedema, swelling, or gum retraction in medical uses, remaining water content in clay for a sculptor, and other quantities familiar to those skilled in an art where the flatscope may be used, our preferred embodiment shows such quantities as annotations visually attached to the operand 419, visible at the same structural point through all motions and changes in viewpoint.

Training system

The images displayed in the flatscope need not be of a current real operand: they may show a recorded procedure. In our preferred embodiment, such a recording stores not only what was shown on the display surface, as a sequence of planar images adjusted to the original operator's position and preferences, but the full input stream from the camera or cameras, and other sources of input. An operator studying such a record cannot modify what is done in the procedure, but can replay it fast or slow, look from different positions or with different zoom factors, modify the apparent lighting, and apply any other data interpretation or analysis subsystem available. This is useful either to a trainee, studying a procedure performed by an expert, or to a mentor or adjudicator assessing skill or problems in performance by the original operator. More actively, the trainee can practice following the tool control of an expert, with both the original tools and the trainee's tools visible in the display: replay speed, magnification factor, zoom and other variables are set either by the trainee or by a trainer or the training system, with repeated or successive procedures adjusted for increasing realism and difficulty. Since the magnification factor and speed affect the difficulty of achieving any particular

motion with the tool tip, this gives a capability to work gradually toward the hardest version, reducing discouragement, improving final results and reducing training time.

Given an adequate system for simulating an operand and its response to tools (simulation of, for example, surgery is an active research field, but means of simulation as such are not the subject of the present disclosure), such a virtual operand can be imported into or connected to a flatscope, allowing a trainee or practitioner to improve without risk to a physical patient or other entity. All the display and interaction elements in the present disclosure apply also in this case of virtual operation.

In either the replayed or virtual modality, the system can add teaching materials to the display, detect errors and report them to the trainee and optionally to a mentor or institutional records.

With either a replayed or simulated procedure, a mentor can be connected from another station, remote or nearby, for the various interactions already discussed.

Flatscopes from hand-held devices

Much of the functionality described above can also be realized by extending the hardware and software of a consumer device such as a personal digital assistant or cameraphone, provided that the included camera is capable of reporting depth (a feature we expect to become common). In our preferred embodiment the device also has means of eye tracking, inbuilt or provided by a supplementary system. In the latter case, the supplementary system is supplied as an integral part of the support frame which makes the device serve this extended function. The voice input features described above require that an application have access to the audio data stream of the device's telephone function, which is typically available, and to sufficient processing power to recognize key words in real time, which will be available in some cases. If the device has built-in voice control, this may be used directly.

The support frame (Drawing 5 [*TBD*]) rests on a work base and holds the device adjustably, as a hands-free intermediate between the operator's eyes and the operand, which is typically at a distance of between 5 and 10 centimeters, though distances moderately outside this range are within the spirit of the invention. The camera included in the device is focussed on the operand, and the resulting view appears on the display screen of the device, usually larger than the original. If the device or the support screen includes means of eye tracking, our preferred embodiment includes an application coded for the device that causes the display to show parallax as described above for the larger system. If the device display includes a means to show images in stereo, our preferred embodiment includes the separate adjustment of the left and right eye images to the actual positions of the left and right eyes.

It is normal for such devices to be network-ready and able to use substantial bandwidth, so all the remote user and mentor functionality described above is applicable to this embodiment also. The key differences are physical scale (the work area and screen are smaller, and the available magnification often less), and control of lighting. Other features such as multispectral imaging can be expected on some devices in this class, but not all, even of those with depth reporting: this limits some functions such as annotation.

Some hand-held devices will have sufficient processing power and adaptability to support flatscope function directly, since programmable multi-core graphics programming units are approaching a physical size that can fit physically within such a

device. Others will not, or will have a graphics/video pipeline that cannot easily be modified to flatscope purposes. However, provided such a device supports fast two-way data transfer by a local connection, the necessary computing power can be included in the base of the support frame. This aspect of the present invention, therefore, ranges from a non-electronic adjustable support frame combined with software running on the hand-held device, to a support frame including a powerful digital system, and software running on the combination of the device and its frame.

Standardization and matching

Doctors often consider the progress of disease and healing by looking at a sequence of photographs, taken over time. However, it is not easy in clinical conditions to arrange the same lighting for every photograph. If a body part is lit in a different direction or at a different distance from the light source, or if the sources vary between daylight and different types of light emitting system, the changes are much harder to assess.

The use of a flatscope for such photography can greatly reduce the problem. First, the light sources 302 can be bright, and dominate the picture, reducing problems with uncontrolled sources such as room lights. Secondly, each image can be automatically uploaded to the patient's electronic medical record (EMR). When any image subsequent to the first is taken, if the sequence is identified, the flatscope downloads the first and overlays it on the current real time image, so that the nurse can easily align the body part with its earlier position for a comparable photograph. Thirdly, in our preferred embodiment the lights 302 are physically arranged to have a range of spectra, so that the record taken on a particular occasion includes measured brightnesses for several different kinds and levels of light. Fourthly, in our preferred embodiment the cameras 301 are multi-spectral in their image response. Fifthly, a standard object with well calibrated reflectivities at many frequencies can be included in each photograph. With such data it is a straightforward problem in linear algebra to construct a least-squares fit to the physical reflectivity of the patient's surface at every visible point, in enough spectral detail to allow a highly accurate reconstruction of what the subject would look like in precisely standardised light, making for easy comparison. Moreover, if the full data set is retained, when an unexpected pathology becomes visible the doctor can revisit the earlier images with a filter adjusted to the spectral properties of different clinical precursors, as well as the current state, with correspondingly improved diagnosis.

A less precise but still useful version of this, with all features but the controlled lighting, is also possible with the hand-held device adaptation described above, even in its form with non-electronic hardware as a support frame, and the software as an application running on the device. The support frame is useful here in helping to standardise the camera position, and freeing the nurse's hands to arrange the patient's position to match the rest of the series.

A more pro-active use of matching applies in the individualization of prosthetics. The latex surface of (for instance) a prosthetic finger should match the appearance of the flesh hand it attaches to, in all probable forms of lighting. Even within any one ethnic group there is considerable variation in the reflective spectrum, which modifies the incident light to produce the reflected light. Lumped by the human visual system into red, green and blue intensities, two such reflected spectra of light originally from (for example) direct sunlight may appear identical, though in each band different frequencies is stimulating the sensors involved. With other incident light (for example,

a fluorscent bulb), the difference is also there, but produces visible difference in red, green and blue response. Adjust material to match in one lighting environment, and there is no guarantee of a match in others. [*add some algebra here*]

The adjustment of latex color consist essentially in adding a mixture of colored materials, with a variety of different reflective spectra when embedded in the latex: different proportions of the light at each different frequency reflected. The final reflectivity is approximately a weighted sum of the spectra of these highly reflective materials (the weights being proportional to their quantity) and the less bright reflection spectrum of the latex itself. Given reflectivity data for the skin in a substantial number of bands within the range of visible light, a least-squares calculation gives the set of weights on the material spectra that most closely combine to match the spectral responses in a standard set of lighting types. Convert those weights into grammes of colorant per gramme of latex, and the result is a material with a good match in all lights to the skin, in its current state. (Unfortunately, the material cannot track the gain or loss of a sun tan, with holidays and seasons.)

System structure

[*add flow charts, etc., as loved by patent examiners: system components, data exchange with remote sites, electronic medical record systesms, etc., etc...*]

What is claimed is:

image gathering and delivery

A1. An imaging system comprising
an image capture apparatus, arranged to capture an image including depth information of an operator work site in communication with a display system;
the display system arranged to receive and display said image including depth information on a display screen to said operator;
wherein said display system is arranged such that the display screen is placed intermediate between the operator's eyes and the hand-sensed position of the work site.

A2. An imaging system according to claim A1, wherein a single screen is intermediate between the eyes of two operators and a shared work site.

A3. An imaging system according to claim A1, wherein a pair of screens is intermediate between the eyes of two operators and a shared work site.

A4. An imaging system according to claim A1, wherein the depth information is obtained by a single camera capable of such sensing.

A5. An imaging system according to claim A1, wherein the depth information is obtained by the use of a plurality of cameras and comparison of the images gathered by said cameras.

A6. An imaging system according to claim A1, wherein the depth information is implicit in the images obtained by the use of a plurality of cameras, and comparison of the images gathered by said cameras occurs in the visual system of the human operator.

A7. An imaging system according to claim A1, wherein the image is obtained by ultrasound emitters and sensors, either mounted on the display system or attached to the subject of the operator's actions, and the display system is enabled to show the image at an operator-perceived position that corresponds to the real position of the scanned subject.

A8. An imaging system according to claim A7, wherein the positioning of the ultrasound emitters and sensors attached to the subject of the operator's actions is tracked and added to the data received by the display system, so that the real position of the scanned subject may be computed.

A9. An imaging system according to claim A1, wherein the image is obtained by sensors detecting radiated or transmitted infrared radiation.

A10. An imaging system according to claim A1, wherein the image is obtained by sensors detecting light in or around the visible range, and classifying it into more than three frequency bands.

A11. An imaging system according to claim A1, wherein the image is obtained by magnetic resonance imaging by a device magnetically and spatially compatible with the system.

A12. An imaging system according to claim A1, wherein images obtained by at least two modalities from a set including but not limited to light, magnetic resonance, ultrasound and infrared radiation are fused to form the image displayed to the operator.

A13. An imaging system according to claim A1, wherein the depth information is conveyed to the operator by stereoscopic management of the views delivered to each eye.

A13a. An imaging system according to claim A1, wherein the system has real time information of the position of one or more of the operator's eyes.

A14. An imaging system according to claim A1, wherein the depth information is conveyed to the operator by parallax as the operator's head moves.

A15. An imaging system according to claim A13a, wherein data specifying the operator's head motion are obtained by a tracking device worn by the operator.

A16. An imaging system according to claim A13a, wherein data specifying the operator's head motion are obtained by one or more cameras mounted on the imaging system and analysis to locate the positions of the operator's eyes.

A17. An imaging system according to claim A1, wherein the image to be displayed for a particular eye is selected from a plurality of camera images.

A18. An imaging system according to claim A1, wherein the image to be displayed for a particular eye is constructed by the use of depth data to adjust the image from a single camera.

A19. An imaging system according to claim A1, wherein the image to be displayed for a particular eye is selected to be the output of a single camera selected from a plurality of differently positioned cameras.

A20. An imaging system according to claim A1, wherein the image to be displayed for a particular eye is synthesized by interpolation between the output of two or more cameras selected from a plurality of differently positioned cameras.

A21. An imaging system according to claim A1, wherein the image to be displayed for a particular eye is synthesized from a three-dimensional surface model derived from the output of two or more cameras selected from a plurality of differently positioned cameras.

A22. An imaging system according to claim A1, wherein the image to be displayed for a particular eye is synthesized from a three-dimensional surface model derived from the output of two or more cameras selected from a plurality of differently positioned cameras.

A23. An imaging system according to claim A1, wherein the display shows transparently, or inserts, supplemental images other than the current contents of the work site, as chosen by the operator for visibility or otherwise at any time.

A24. An imaging system according to claim A23, wherein the supplemental images include recorded replays of the operator's actions, at standard or modified speed.

A25. An imaging system according to claim A23, wherein the supplemental images include data from records of the subject of the operator's actions.

A26. An imaging system according to claim A23, wherein the supplemental images include instructional material relevant to the operator's task.

A27. An imaging system according to claim A1, wherein the image data are made available to external communication systems, for storage or for display, following industry standards for interoperability and data exchange.

A28. An imaging system according to claim A23, wherein an external agent may annotate the images displayed.

A29. An imaging system according to claim A23, wherein the supplemental images include data from annotation by an external agent.

A30. An imaging system according to claim A23, wherein the supplemental images include data from scans or images created at an earlier time.

A31. An imaging system according to claim A30, wherein the scans or images created at an earlier time are warped to match the current shape of the subject of the operator's actions.

A32. An imaging system according to claim A23, wherein the supplemental images include data from annotation by automatic image analysis.

A33. An imaging system according to claim A 32, wherein the supplemental images move in the display to retain the same apparent location as the subject of the operator's actions.

A34. An imaging system according to claim A23, wherein the automatic image analysis finds distances, lengths, widths, or areas in the subject of the operator's actions.

A35. An imaging system according to claim A23, wherein the automatic image analysis analyses surface shape in the subject of the operator's actions.

A36. An imaging system according to claim A23, wherein the automatic image analysis analyses surface texture in the subject of the operator's actions.

A37. An imaging system according to claim A23, wherein the automatic image analysis analyses colour in a plurality of frequency bands in the subject of the operator's actions.

A38. An imaging system according to claim A27, wherein the automatic image analysis analyses temperature in the subject of the operator's actions.

A39. An imaging system according to claim A27, wherein the automatic image analysis analyses the time course of a change in properties caused by the operator's actions.

A40. An imaging system according to claim A27, wherein the automatic image analysis derives material or biomedical properties the subject of the operator's actions.

A41. An imaging system according to claim A32 to A40, wherein the annotations report changes relative to measures computed earlier.

A42. An imaging system according to claim A32 to A41, wherein the automatic analysis integrates data from a plurality of the modalities there specified, with optionally data from other modalities.

A43. An imaging system according to claim A23 to A42, wherein the annotations report measures of surface elasticity, necrosis, temperature, vascularisation, inflammation, infection, perfusion, oedema, swelling, gum retraction, ...

control widgets

A44. An imaging system according to claim A1, wherein the detected positions of the tools controlled by the operator provide input to the system.

A45. An imaging system according to claim A44, wherein the detected positions are used to select modes of input.

A46. An imaging system according to claim A44, wherein the detected positions are used to select points or substructures in the displayed object.

A47. An imaging system according to claim A44, wherein the detected positions are used to mark annotations on the displayed object.

A48. An imaging system according to claim A44, wherein the detected positions are used to define spatial forms in the space of the displayed object.

A49. An imaging system according to claim A44, wherein the open or closed state of the tool tip provides input for the grasping or release of a virtual object.

A50. An imaging system according to claim A44, wherein the motion of the tool tip is used to manipulate the shape of a virtual object.

A51. An imaging system according to claim A49, wherein the motion and state of the tool tip are used to manipulate the shape of a virtual object.

A52. An imaging system according to claim A13a, wherein the eye positions are used to control zoom of all or part of the displayed image.

training system

B1. An interaction system comprising
a motion capture apparatus, arranged to capture the motion of manipulation tools by an operator, in communication with a display system;
the display system arranged to receive and display said motion, transferred to computer generated tool images, to said operator;
the display system further arranged to display additional images;
the display system further arranged to record and replay events;
wherein said display system is arranged such that the display screen is placed intermediate between the operator's eyes and the hand-sensed position of the work site.

B2. An interaction system according to claim B1, wherein the additional images are of virtual objects that can be manipulated by the operator.

B3. An interaction system according to claim B1, wherein the scale relation between the captured motion and the transferred motion is adjustable by the operator.

B4. An interaction system according to claim B1, wherein the operator selects that the replay of recent events is faster or slower than real time.

B5. An interaction system according to claim B1, wherein the additional images are of a recorded procedure performed on real or virtual objects by another operator.

B6. An interaction system according to claim B1, wherein the system reports how well the motions of the current operator conform to the motions of the recorded operator.

B7. An interaction system according to claim B1, wherein the additional images are of training materials, textual information, or a mentor.

B8. An interaction system according to claim B1, wherein a mentor operates another display and can communicate by voice or video.

B9. An interaction system according to claim B8, wherein the mentor can impose motion on virtual tools active and visible within the display.

B10. An interaction system according to claim B9, wherein the system reports visually and quantitatively the operator's success in matching the motion of the mentor's tools, recorded or controlled by current input from the mentor.

eye-tracking applications

C1. An interaction system comprising
a display screen;
one or more cameras directed at the operator;
a computer equipped with algorithms to identify the position and direction of the eyes of the operator, to recognise steady gaze at a point within the display screen, and to respond to head motion by zooming or shrinking the displayed image.

C2. An interaction system according to claim C1, wherein the display screen shows images stereographically, by showing different images to each eye.

C3. An interaction system according to claim C1, wherein the display screen shows images with parallax, by adjusting the image according to the position of the eyes separately, or a single intermediate position.

C4. An interaction system according to claim C1, wherein the zoom or shrinking keeps fixed the point at which gaze is steady.

C5. An interaction system according to claim C4, wherein the zoom or shrinking is uniform.

C6. An interaction system according to claim C4, wherein the zoom or shrinking is greater in the vicinity of the fixed point, and the distant surroundings move less or not at all.

C7. An interaction system according to claim C1, wherein the zoom or shrinking applies to a planar image.

C8. An interaction system according to claim C1, wherein the zoom or shrinking applies to a three-dimensional image, and the apparent depth of the fixed point relative to the screen remains fixed while the apparent sizes of the objects around it change.

PDA/cameraphone/camera as manipulation magnifier

D1. An interaction system comprising
a device of hand-held size, equipped on one side with a video camera that can focus in the near field, and on the other a display screen;
an adjustable stand for the device, by which it may be placed between the eyes and the hands of an operator;
a software application which brings to the display screen the operator's hands and any objects manipulated by them, at a size and focus depth controllable by the operator.

D2. An interaction system according to claim D1, wherein the operator controls the sensitivity, dynamic range and spectral characteristics of the camera, and the manner in which the camera data are represented in the display.

D3. An interaction system according to claim D1, wherein the device includes also a camera directed toward the operator, and control is by tracking of eye movements.

D4. An interaction system according to claim D1, wherein the device includes also a microphone directed toward the operator, and control is by voice.

D5. An interaction system according to claim D1, wherein the device includes also a network connection, and other persons equipped with any net-connected display can watch the same view as the operator.

D6. An interaction system according to claim D4, wherein a person watching remotely can control the view.

D7. An interaction system according to claim D4, wherein a person watching remotely can by voice or image advise the operator.

D8. An interaction system according to claim D2, wherein an application allows the operator to choose spectral representation that is particularly indicative of a particular state of the object being manipulated.

D9. An interaction system according to claim D8, wherein the particular state is temperature, vascularisation, inflammation, infection, perfusion, oedema, swelling, gum retraction, ...

D10. An interaction system according to claim D8 or D9, wherein the particular state is detected by algorithmic analysis and shown by colour coding.

Lighting standardiser snd color matcher

E1. An interaction system comprising
a camera capable of detecting light in or around the visible range, and classifying it into more than three frequency bands;
a means of transferring data to a permanent store;
a computing device equipped with an analysis algorithm capable of determining from the detected light the lighting conditions and spectral reflectivities of a photographed object, and deducing therefrom an image in at least three frequency bands that shows the image approximately as it would have been with a user-specified type and brightness of illumination;
a means to attach the resulting image by date and identity to a time series of images of the same subject.

E2. An interaction system according to claim E1, wherein the analysis algorithm is supported by the inclusion in the image of a standardised object whose reflectivity is known for all the frequency bands supported by the camera.

E3. An interaction system according to claim E1, wherein the more than three frequency bands include the near infrared.

E4. An interaction system according to claim E1, wherein the more than three frequency bands include more than one band within each of the standard sensitivity bands, red, green and blue, of the human optical system.

E5. An interaction system according to claim E1, wherein a further algorithm matches the multi-spectral reflectivity of a region of the image, as computed by the image processing system to a mixture of materials with known reflectivity, that will closely reproduce under many lighting conditions the colours observed by the human optical system when observing the material corresponding to that region of the image.