Recent Advances and New Applications of Computer Vision

Roberto Cipolla Carlos Hernández George Vogiatzis Björn Stenger

Computer Vision Group, Cambridge Research Laboratory, Toshiba Research Europe Ltd.
Department of Engineering, Cambridge Univ.

1 Introduction - the 3Rs of Computer Vision

Computer vision technology is beginning to find a place in a number of consumer products including camera phones; interfaces to games consoles; assisting parking and driving in automobiles; image and video search on a computer and the internet and more recently internet shopping.

At Cambridge we have identified a number of possible applications and are now pioneering core technology in the main three areas of vision - Reconstruction (3D shape recovery from uncalibrated images); Registration (human body detection and tracking for use in novel interfaces) and Recognition (object detection, segmentation and recognition in video). Our approach is predominantly geometric but includes modern practice in machine learning.

In the following article we briefly review our research in two areas: (1) 3D shape recovery and (2) simple and robust interfaces to computers by detecting and tracking hands.

2 Reconstruction of Shape

We begin by examining the problem of obtaining a complete, detailed model of a real-world 3D object, given a sequence of images of that object. This topic has been studied extensively since the earliest days of machine vision (e.g. [7]) by researchers aiming to understand the human visual system through construction of computer algorithms. In recent years, due to the dramatic improvement in computational power as well as the increased availability of digital imaging technology, reconstruction of shape from images has received interest as a practical application.

Accurate geometric models that can be used to synthesise realistic novel views of the objects (see **Figure 1**) are highly desirable. The most common ways of obtaining such models are either by manually constructing them in a CAD program, or by using laser range scanning technology (e.g. [5]). The manual method is quite impractical and error-prone for large scale, complex models, while laser range scanning and other similar techniques remain prohibitively expensive for a wide range of potential applications. Consequently, the automatic acquisition

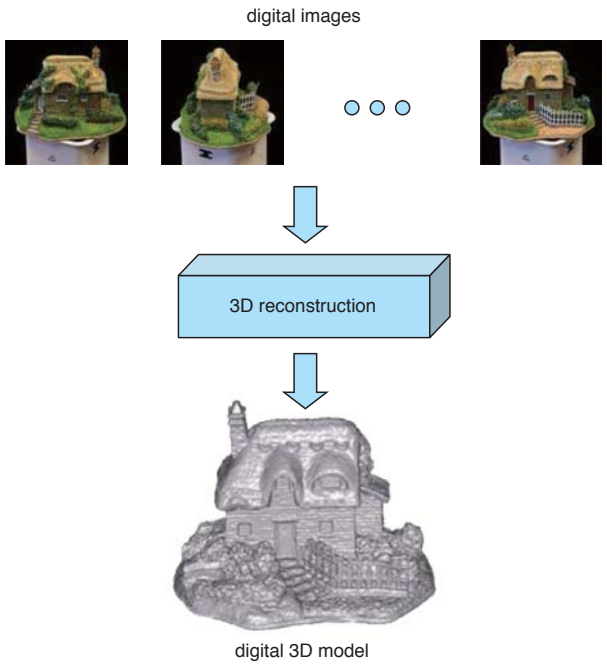

Figure 1: Visual estimation of shape. Our technologies enable the user to obtain a complete, detailed, three-dimensional model of a real object from a collection of digital photographs. This accurate digital copy can then be used to inspect details of the object from any viewpoint, measure geometric properties such as angles, lengths and volumes or reproduce it in a different material.

of photo-realistic 3D models from digital images of the scene emerges as a cheap, lightweight and non-intrusive alternative which has already found applications in archaeology [9], modelling of architecture [1] and digitisation of sculpture [4] among others.

2.1 Challenges

Despite the optimism of early Computer Vision researchers, a fully automated Visual Reconstruction system remains elusive [3]. Some of the key difficulties, adapted here from [11], are the following:

High dimensionality Representing a general scene's geometry and reflectance requires infinitely many degrees of freedom. Estimating those unknowns is generally infeasible unless strong priors about both geometry and reflectance are applied.

Photometric ambiguity The observed intensity of a pixel depends on the surface geometry at the corresponding scene point, its local reflectance as well as light in a non trivial way. From that intensity these properties can be constrained but cannot be directly estimated.

Loss of depth Camera images of a scene are formed by projecting 3D space to a 2D plane. During this process the distance travelled by light between scene and camera (i.e. depth) is lost. Although there are situations where this ambiguity can be resolved in the monocular case as in Shape from Shading, human and artificial vision systems alike typically employ multiple images of the scene from varying viewpoints and/or under varying illumination.

We have developed two different approaches to the visual shape reconstruction problem, each of which is particularly suited for a different class of solid objects. Our first method can handle objects made of richly textured materials while the second method deals with completely untextured objects such as white porcelain statues.

2.2 Textured materials

The estimation of shape from image correspondences, sometimes referred to simply as *dense stereo* is a very powerful technique for reconstructing a scene given M images of this scene from *different* viewpoints. It is based on the following very simple observation: A 3D point located *on* the scene surface projects to image regions of *similar* appearance in all images where it is not occluded. Equivalently, this principle can be stated in terms of visual rays. As mentioned above, each image location corresponds to a 3D line. Given a number of image locations that depict the same scene location, the intersection of their visual rays [1] will be that scene location. This is illustrated in **Figure 2**.

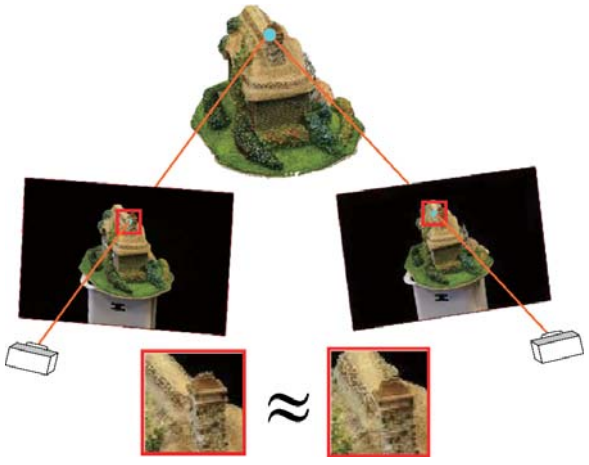

Figure 2: Shape from image matches. The basic principle underlying our algorithm for reconstructing textured objects is the following: Given two images in which the camera position is known, if two pixel positions correspond to the same object in 3D space (e.g. the roof of the house) then by triangulating from the camera centres one can compute the exact 3D location of that object. Conversely, to examine if a particular 3D location is on the surface of the object we can project that location to the two images, and compare the nearby regions. If they are similar the 3D point is on the surface.

Most work in the dense stereo problem assumes a Lambertian reflectance model for the surface as well as constant illumination throughout the sequence. These conditions imply that a scene point projects to pixels of the same intensity in images where it is visible, which makes the task of identifying matching

Figure 3: Toy House. This is an example of a 3D model of a real object, obtained using our technology. In the top row are four images of a toy house while in the bottom row, renderings of the 3D model from similar viewpoints are shown . The first three images were part of the input sequence used while the fourth was not shown to the algorithm. The model of this small toy house (approximately 10cm in diameter) contains accurately reconstructed sub-millimetre details such as the fence and the relief of the roof.

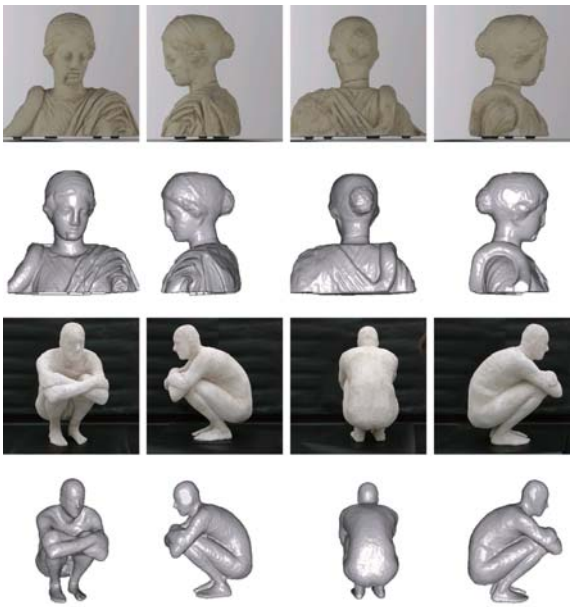

Figure 4: Digitising sculpture. The first two rows show the reconstruction of a plaster bust of the Greek goddess Hygeia. The input sequence consists of 36 images. Four of these are shown in the first row while the second row shows similar views of the reconstructed model. The third and fourth are input images and reconstructed model for a *crouching man* sculpture by the British sculptor Anthony Gormley. The input sequence in that case consisted of 72 images.

[1] If there is motion of the camera centre for at least two images, the visual rays will not all be coincident and will therefore have a well defined intersection point.

image locations easier. Additionally it is assumed that the object is well textured so that parts of the object surface can be uniquely identified in multiple images.

We have developed a volumetric formulation for the 3D reconstruction problem which is amenable to a computationally tractable global optimisation using Graph-cuts. Our approach is to seek the optimal partitioning of 3D space into two regions labelled as 'object' and 'empty' under a cost functional consisting of the following two terms: (1) A term that forces the boundary between the two regions to pass through photo-consistent locations and (2) a ballooning term that inflates the 'object' region. To take account of the effect of occlusion on the first term we use an occlusion robust photo-consistency metric based on Normalised Cross Correlation, which does not assume any geometric knowledge about the reconstructed object. The globally optimal 3D partitioning can be obtained as the minimum cut solution of a weighted graph. Some of the objects we have digitised using this technique can be seen in **Figures 3** and **4**.

3 Photometric Stereo and untextured materials

While dense stereo techniques offer detailed full 3D reconstructions, they rely on richly textured objects to obtain correspondences between locations in multiple images which are triangulated to obtain shape. As a result these methods are not directly applicable to the class of completely untextured objects due to the lack of detectable surface features. On the other hand, photometric stereo works by observing the changes in image intensity of points on the object surface as illumination varies. These changes reveal the local surface orientations at those points that, when integrated, provide the 3D shape. Because photometric stereo performs integration to recover depth, much less regularisation is needed and results are generally more detailed. Furthermore, photometric stereo makes fewer assumptions about surface texture and reflectance, which can be almost completely arbitrary as demonstrated in [2]. However, the simplest way to collect intensities of the *same* point of the surface in multiple images is if the camera viewpoint is held constant, in which case every pixel always corresponds to the same point of the surface. This is a major limiting factor of the method because it does not allow the recovery of the full 3D geometry of a complex many-sided object such as a sculpture. Due to this limitation existing photometric stereo techniques have so far only been able to extract depth-maps (e.g. [10]) with the notable recent exceptions of [12, 6], where the authors present techniques for recovering 2.5D reconstructions from multiple viewpoints. The full reconstruction of many-sided objects is however still not possible by these methods. While in theory one could apply photometric stereo from multiple viewpoints and then merge the multiple depth-maps of the object into a single 3D representation, in practice this procedure can be complicated and error-prone.

We have developed a different solution to this problem by exploiting the powerful silhouette cue. We modify classic photometric stereo and cast it in a multi-view framework where the camera is allowed to circumnavigate the object and illumination is allowed to vary. The setup for this technique is illustrated schematically in **figure 5**. Firstly, the object's silhouettes are used to recover camera motion using a technique similar to [8], and via a novel robust estimation scheme they allow us to accurately estimate the light directions and intensities in every image.

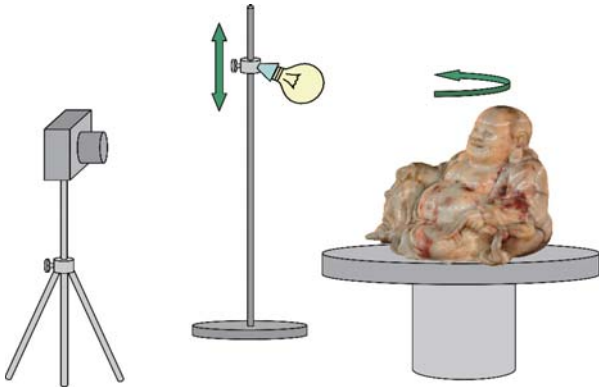

Figure 5: Acquisition Setup for textureless objects. The object is rotated on a turntable in front of a camera and a point light-source. A sequence of images is captured while the light-source changes position between consecutive frames. No knowledge of the camera or light-source positions is assumed.

Secondly, the object surface, which is parameterised by a mesh and initialised from the visual hull, is evolved until its

Figure 6: Reconstructing porcelain vases. Two Chinese porcelain vases (Qing dynasty, Qianlong period, 1736-1795) with very fine relief are reconstructed from two sequences of 36 images each. Top: sample of input images. Bottom: our reconstruction results. The surface captures all the fine details present in the images, even in the presence of strong highlights.

Figure 7: Reconstructing coloured marble. A marble Buddha figurine (Chinese, Qing dynasty) is reconstructed from a sequence of 36 images.

predicted appearance matches the captured images. Each face of the mesh is projected in the images where it is visible and the intensities are collected. From these intensities and the illumination computed previously, a normal direction is assigned to each face by solving a local least squares problem. The mesh is then iteratively evolved until these directions converge to the actual surface normals of the mesh. These two phases are then repeated until the mesh converges to the true surface. The advantages of our approach are the following:

- It is fully uncalibrated: no light or camera pose calibration object needs to be present in the scene.
- The full 3D geometry of a complex, shiny, textureless object is accurately recovered, something not previously possible by any other method.
- It is practical and efficient as evidenced by our simple acquisition setup.

Figures 6 and **7** show some digitised artefacts from the collection of the Fitzwilliam museum in Cambridge.

4 Reconstruction of Deforming Shape

Even though the previous section shows how advanced is state-of-the-art on 3D modelling of rigid objects, very little work has been accomplished on non-rigid scenes so far. Modelling deforming objects is extremely important for tasks such as cloth reconstruction but also for general dynamic surface modelling, such as body or faces. In this section we explore the associated capture methodology to acquire the detailed 3D shape, bends, and wrinkles of deforming surfaces. Moving 3D data has been difficult to obtain by methods that relied on known surface features, structured light, or silhouettes.

Multispectral photometric stereo is an attractive alternative because it can recover a dense normal field from an un-textured surface. Experiments were performed on video sequences of un-textured cloth, filmed under spatially separated red, green, and blue light sources.

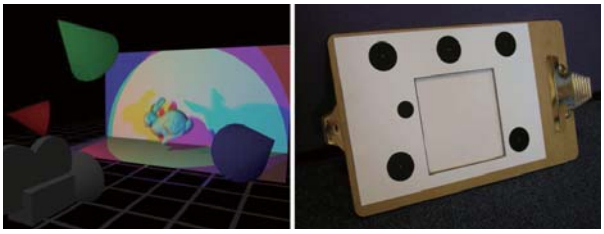

Figure 8: Setup and calibration board. Left: A schematic representation of our multispectral setup. Right: Linking the two clipboards and attaching a printed calibration pattern produces a planar trackable target for computing the orientation of the pattern plane. Cloth inserted between the boards appears in the square hole and creates the association between colour and orientation.

The proposed technique for acquiring complex motion data from real moving cloth, uses a practical setup that consists of an ordinary video camera and three coloured light sources (see **Figure 8**). The key observation is that in an environment where red, green, and blue light is emitted from different directions, a Lambertian surface will reflect each of those colours simultaneously without any mixing of the frequencies. The quantities of red, green and blue light reflected are a linear function of the surface normal direction. A colour camera can measure these quantities, from which an estimate of the surface normal direction can be obtained. By applying this technique to a video sequence of a deforming object one can obtain a sequence of normal maps for that object and integrate them to produce a sequence of depth-maps (see **Figure 9**).

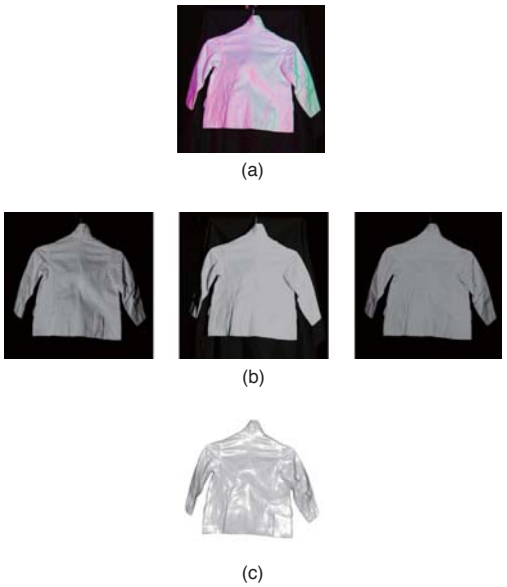

Figure 9: photometric stereo with coloured light. (a) Single frame from a jacket sequence, where the same object is illuminated *simultaneously* by three different coloured lights. (b) Equivalent three grayscale images, each corresponding to a different illumination, providing the input to a classic photometric stereo reconstruction [17] shown in (c).

4.1 Depth-map video

In this section we follow the exposition of Kontsevic *et al.*[15]. For simplicity, we first focus on the case of a single distant light source with direction **l** illuminating a Lambertian surface point P with surface orientation direction **n**. Let $S(\lambda)$ be the energy distribution of that light-source as a function of wavelength λ and let $\rho(\lambda)$ be the spectral reflectance function representing the reflectance properties at that surface point. We assume our camera pixels consist of three sensors sensitive to different parts of the spectrum. If $\nu_i(\lambda)$ is the spectral sensitivity of the i-th sensor for the pixel that receives light from P, then intensity measured at that sensor is

$$r_i = \mathbf{l} \cdot \mathbf{n} \int S(\lambda)\rho(\lambda)\nu_i(\lambda)d\lambda \qquad (1)$$

or in matrix form

$$\mathbf{r} = M\mathbf{n} \qquad (2)$$

where the (i, j)th element of M is

$$m_{ij} = l_j \int S(\lambda)\rho(\lambda)\nu_i(\lambda)d\lambda. \qquad (3)$$

When more light sources are added, if the system is linear and $\mathbf{l} \cdot \mathbf{n} \geq 0$ still holds for each light, the response of each sensor is just a sum of the responses for each light source individually, leading to equation (2) still being valid with:

$$M = \sum_k M^k \qquad (4)$$

where M^k describes the k-th light source. Since each of the M^k is of rank 1, this implies that in the absence of self occlusions, a minimum of three different lights needs to be present in the scene for M to be invertible. If the surface is uniformly coloured, then $\rho(\lambda)$ and consequently M will be constant across all unoccluded locations.

Equation (2) establishes a 1-1 mapping between an RGB pixel measurement from a colour camera and the surface orientation at the point projecting to that pixel. Our strategy is to use the inverse of this mapping to convert a video of a deformable surface into a sequence of normal maps. We estimate the 1-1 mapping by employing an "easy-to-use" calibration tool (figure 8 (left)). The pattern is planar with special markings that allow the plane orientation to be estimated. By placing the object in the centre of the pattern, we can measure the colour it reflects at its current orientation. We thus obtain a sequence of (**r**,**n**) pairs to which we fit the mapping M using linear least squares.

4.2 Depth from Normals

By estimating and then inverting the linear mapping M linking RGB values to surface normals, we can convert a video sequence captured under coloured light into a *video of normal-maps*. Due to the dark room conditions, by simple intensity thresholding we can segment background pixels in every frame of the original video, as they are almost perfectly black. We then integrate each normal map independently to obtain a depth map in every frame by imposing that the occluding contour is always at zero depth. This integration process is a fairly established technique and several algorithms are available. We have used the Successive Overrelaxation solver (SOR) [13] because of its robustness and simplicity. At the end of the integration process, we obtain a *video of depth-maps*.

In **Figure 10** we show several views of frame 380 without the texture map in high resolution (the mesh consists of approximately 180,000 vertices). The images clearly show the high frequency detail of the sweater. To the best of our knowledge, this is the only method able to reconstruct deforming cloth with such detail.

Figure 10: Cloth reconstruction results of a deforming sweater. Multispectral photometric reconstruction of a single frame of a longer video sequence using the technique described in section 4.1. From left to right, multiple viewing angles of frame 380 of the sweater sequence.

To demonstrate the potential of our method for capturing cloth for animation, we captured a sequence of moving cloth, registered it using optical flow [14], and attached the registered meshes to an articulated skeleton. Skinning algorithms have varying degrees of realism and complexity, e.g. [16]. **Figure 11** shows example frames from the rendered sequence. Even though the skeleton and cloth motions are not explicitly aligned, the visual effect of the cloth moving on a controllable character is appealing. Such data-driven cloth animation can serve as a useful tool and presents an alternative to physical cloth simulation.

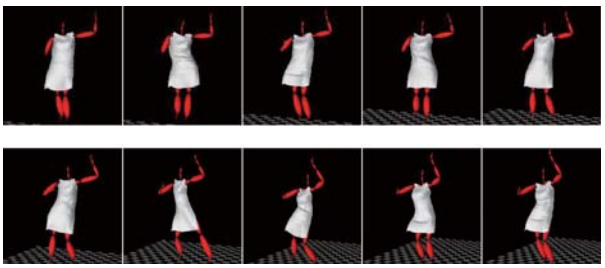

Figure 11: Attaching captured moving cloth to an animated character. We apply smooth skinning to attach a moving mesh to an articulated skeleton that can be animated with mocap data. The mesh is simply animated by playing back one of our captured and registered cloth sequences, in this case a dancing sequence.

4.3 Summary

Building on the long established but surprisingly overlooked theory of multispectral lighting for photometric stereo, we have discovered and overcome several new obstacles. We developed a capture methodology that parallels existing work for capturing static cloth, but also enables one to capture the changing shape of cloth in motion. Integration of the resulting normal fields is already possible with the simple boundary condition that the occluding contour is at zero depth. We have verified the accuracy of the depth-maps against classic photometric stereo. When such a sequence of surfaces is played back, it appears to be changing smoothly. The high level of detail captured by the normal fields includes surface bends, wrinkles, and even temporary folds.

Finally, with access to a unique stream of rich 3D cloth poses, we have shown how easily the data is employed in a creative context for realistic character animation of a clothed avatar.

5 Gesture User Interfaces Using Computer Vision

Computer vision allows touch-free input via hand gestures. Together with the Multimedia Laboratory we are pursuing research into making such methods robust, efficient and intuitive.

A mouse and keyboard are the standard input devices for most personal computers. The mouse as a pointing device has a number of limitations. First of all, for some applications it may be beneficial to have more degrees of freedom than two provided by the mouse input. It can also be practical to use both hands to control certain applications, such as manipulating (e.g. stretching) objects on the screen. From an ergonomic point of view the switching between keyboard and mouse input can be cumbersome and repeated mouse use can lead to stress injuries. Also, it is not a natural input device for handwriting and drawing applications. Designers use drawing pads for such tasks, which are precise, but are also an extra investment and have an initial learning curve. Both mouse and drawing pad also require extra desk space. However, some recent computers ship with a CCD camera attached or already integrated to facilitate video calls or conferencing applications.

5.1 A Pointing Interface By Detecting Changes of Image Topology

This section presents an intuitive visual pointing interface for custom personal computers, where a camera is mounted on top of the screen and is pointed at the keyboard. The user can easily switch between keyboard and gesture input. Further, the gesture input mode has more degrees of freedom than a conventional computer mouse. Thus it can provide a natural interface for certain tasks such as hand writing, image manipulation or visual navigation.

The hand is seen as a foreground object, which is separated from the background by colour segmentation. The method uses a connected components algorithm that detects changes in the shape topology of the background. The insight here is that by touching two finger tips, e.g. by touching index finger and thumb, the number of background segments increases by one. By fitting an ellipse to this new shape one can determine location, orientation and scale of the hand. Additionally, one can estimate the hand direction, thus controlling a point that is close to the finger tips. **Figure 12** shows the system setup.

Figure 12: System setup. A camera is mounted on top of the monitor and is directed at the user's hands. A pointing device can be implemented by foreground segmentation and shape analysis.

Shape analysis As a first processing step the system performs foreground segmentation. This is done by initialising a colour model using a boosted hand detector [19, 20]. Skin colour is extracted from an ellipse in the centre of the detected region and adjacent regions are taken as background regions, see **Figure 13**. The resulting distributions are then used to compute the skin colour likelihood for each pixel. These values can serve as an input for any binary segmentation algorithm. Hysteresis thresholding followed by a morphological opening and closing operation are performed to remove misclassification due to pixel noise. A typical result can be seen on the right of Figure 13. Note that the hand's reflection in the screen as well as the wooden table leads to some mis-classified pixels.

The next step is to analyse the shape of the foreground map. For this we assume that the background region normally consists of a single connected component. When the foreground forms a loop, for example, when the thumb and the index finger form a round shape this is no longer the case. This principle is used here

Figure 13: Colour model initialisation. Left: An open and flat hand is located using a boosted detector. The skin colour distribution is learned from the values in the centre of the detected region. Middle: The resulting likelihoods for skin colour, Right: A post-processed binary map.

to implement a point and click interface. A connected component algorithm is used to count the number of connected segments. When there are more than two segments, these are used as input regions for the pointing interface. First we fit an ellipse to each such region. The four ellipse parameters can directly be used as input to the interface. In addition to the two parameters of a conventional mouse, this includes orientation and scale.

Finger tip detection A complementary shape analysis technique is to detect finger tip regions. Here we present a simple algorithm which is efficient but requires a relatively clean foreground segmentation. We first compute the foreground contour and assume that points corresponding to a finger tip lie on this contour. We then compute the foreground/background ratio in a local neighbourhood around all contour points and only keep those whose ratio is between two threshold values. Among multiple candidate detections around one finger tip candidate the one with the smallest ratio is chosen as output location. Advantages of this method are that (a) it is very fast to evaluate and (b) it is relatively tolerant to a noisy contour shape, e.g. caused by shadows. On the other hand, the method does require a rough global scale estimate in order to determine the local neighbourhood in which to compute the ratio values.

Experimental results The finger tip detection algorithm is very efficient and has potential for various input modes. Illustrative examples are shown in **Figure 14**. Individual finger tips can be detected, however when fingers touch each other, thus changing the foreground contour shape, they are no longer detected.

The proposed algorithm based on image topology runs in real-time. We show a couple of illustrative results of the pointing algorithm. **Figure 15** shows some frames from an example sequence using one or both hands for pointing. It can be used analogously to a mouse pointer, although currently no cursor is being moved when the fingers are not creating a closed loop. In addition to the location the orientation and scale can be detected reliably. A simple drawing application is shown in **Figure 16**. The hand can be moved like a holding a pen. When the fingers are put together the drawing mode is activated and a line segment is drawn on the screen.

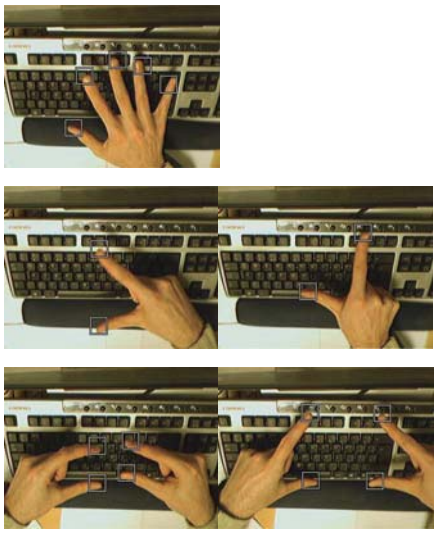

Figure 14: Finger tip detection examples. (row 1) All fingers can be detected simultaneously, however only when the points lie on a contour and part of the finger contour is visible, (row 2) hand rotation can be used to accurately describe 2D rotations, (row 3) scaling operations can be executed by dragging the fingers.

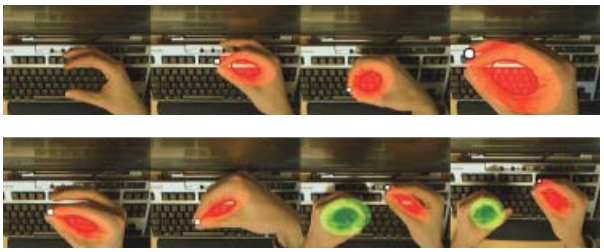

Figure 15: Example sequence. This sequence shows frames of an input sequence. The pointer is activated by forming a closed loop with thumb and index finger. This way one or more regions can be defined whose location, orientation and scale can be used as input to the interface.

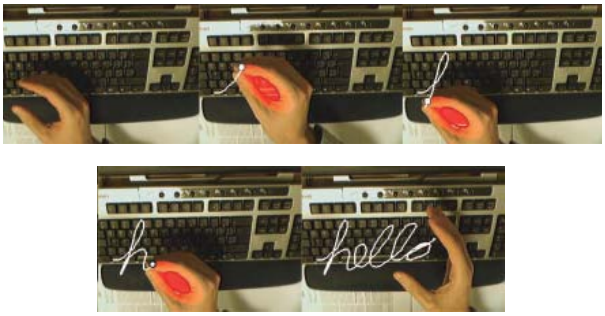

Figure 16: Writing application. Writing and drawing is natural with the proposed interface.

The advantages of the system are the ease with which the user can switch between the typing and pointing as well as additional degrees of freedom through scale and orientation estimation. Further, multiple detections can be handled seamlessly. This input method may be able to replace or drawing pad use in some cases. Furthermore, finger tip detection also has potential for accurate input, however some issues need to be resolved first: detection reliability needs to be improved and it should be combined with the detection fingers touching each other. This will be a topic of further research.

6 Conclusions

We have very briefly outlined 4 of the projects being carried out in Cambridge. In the area of 3D shape recovery from uncalibrated images in controlled indoor environments we have made significant progress. The next major challenge is to be able to

build models of the outdoor world with less control over lighting and background and to make these algorithms more efficient.

In the area of hand detection and tracking more research is needed in making the algorithms more reliable. The existing technology developed at Cambridge and the Multimedia Laboratory is now ripe for exploitation in simple gesture interfaces to computers and televisions. However, although the core technology is nearly ready, a lot of careful HCI [2] design is needed to make interfaces that are simple, intuitive, useful and fun to use!

References

[1] A. Dick, P. H. S. Torr, S. Ruffle, and R. Cipolla. Combining single-view recognition and multiple-view stereo for architectural scenes. In *Proc. 8th Intl. Conf. on Computer Vision*, pages 268–274, 2001.

[2] D. B. Goldman, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and spatially-varying brdfs from photometric stereo. In *Proc. 10th Intl. Conf. on Computer Vision*, 2005.

[3] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2004.

[4] C. Hernández and F. Schmitt. Silhouette and stereo fusion for 3d object modeling. *Computer Vision and Image Understanding*, 96(3):367–392, December 2004.

[5] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk. The digital michelangelo project: 3d scanning of large statues. In *Proc. of the ACM SIGGRAPH*, page 1522, 2000.

[6] J. Lim, J. Ho, M. Yang, and D. Kriegman. Passive photometric stereo from motion. In *Proc. 10th Intl. Conf. on Computer Vision*, 2005.

[7] D. Marr. *Vision*. W.H.Freeman & Co., 1982.

[8] P. R. S. Mendonç.a, K.-Y. K. Wong, and R. Cipolla. Epipolar geometry from profiles under circular motion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(6):604– 616, 2001.

[9] M. Pollefeys, R. Koch, Vergauwen M., and L. Van Gool. Metric 3d surface reconstruction from uncalibrated image sequences. In Springer-Verlag, editor, *Proceedings of SMILE Workshop (post-ECCV'98)*, pages 138–153, 1998.

[10] A. Treuille, A. Hertzmann, and S. Seitz. Example-based stereo with general brdfs. In *Proc. 8th Europ. Conf. on Computer Vision*, may 2004.

[11] M. Weber. *Curve and Surface Reconstruction from Images and Sparse Finite Element Level-Sets*. PhD thesis, Cambridge University, 2004. PhD Thesis.

[12] L. Zhang, B. Curless, A. Hertzmann, and S. Seitz. Shape and motion under varying illumniation: Unifying structure from motion, photometric stereo, and multiview stereo. In *Proc. 9th Intl. Conf. on Computer Vision*, 2003.

[13] M. E. Davis and J. A. McCammon. Solving the finite difference linearized poisson-boltzmann equation: A comparison of relaxation and conjugate gradient methods. *J. Comput. Chem.,* 10(3):386–391, 1989.

[14] C. Hernández, G. Vogiatzis, G. Brostow, B. Stenger, and R. Cipolla. Non-rigid photometric stereo with coloured light. In *to appear in ICCV*, 2007.

[15] L. L. Kontsevich, A. P. Petrov, and I. S. Vergelskaya. Reconstruction of shape from shading in color images. *J. Opt. Soc. Am. A*, 11(3):1047–1052, 1994.

[16] J. P. Lewis, M. Cordner, and N. Fong. Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 165–172, 2000.

[17] R.J. Woodham. Photometric method for determining surface orientation from multiple images. In *Optical Eng.,* number 1, pages 139–144, 1980.

[18] T. Ike and B. Stenger. A real-time hand gesture interface implemented on a multicore processor. In *Machine Vision Applications*, Tokyo, Japan, May 2007.

[19] T. Mita, T. Kaneko, B. Stenger, and O. Hori. Discriminative feature co-occurrence selection for object detection. *IEEE Trans. Pattern Analysis and Machine Intell.,* 2008. to appear.

[20] B. Stenger. Template-based hand pose recognition using multiple cues. In *Proc. ACCV*, pages 551–560, Hyderabad, India, January 2006.

[2] HCI: Human Computer Interface

Prof. Dr. Roberto Cipolla
Managing Director
Cambridge Research Laboratory,Toshiba Research Europe Ltd.
Professor of Information Engineering, Cambridge Univ.

Dr. Carlos Hernández
Research Scientist
Computer Vision Group, Cambridge Research Laboratory,
Toshiba Research Europe Ltd.

Dr. George Vogiatzis
Research Scientist
Computer Vision Group, Cambridge Research Laboratory,
Toshiba Research Europe Ltd.

Dr. Björn Stenger
Research Scientist
Computer Vision Group, Cambridge Research Laboratory,
Toshiba Research Europe Ltd.