

Processes for Freehand Image Capture: HP CapShare Technology

*Ross R. Allen, Raymond G. Beausoleil
Hewlett-Packard Laboratories
Palo Alto, California USA*

Abstract

The HP CapShare 920 information appliance is an example of a new generation of self-contained devices developed to enable the capture of text and images with a freehand swipe over a document. This process is fundamentally different from conventional hand-held scanners, which require careful attention to scan velocity, a linear path, and a PC to assemble the image from successive swaths. The HP CapShare 920 is entirely optoelectronic with a linear array of photodetectors that is free to translate and rotate in three planar degrees of freedom during image capture at velocities up to 16 inches per second. Novel "navigator" ICs determine the velocity of the end-points of the array to reconstruct the path. The reconstruction of the scanned image, including compensation for nonlinear motion and swath stitching, is done entirely within the appliance and does not rely on a host PC. During image capture, about 2.8 billion operations are performed each second.

electromechanical motion encoders, often rollers or wheels with friction coatings, are used to measure displacement and velocity during image capture and to constrain motion — usually to a straight line. Recently, scanners have appeared with roller ball encoders that allow more freedom of motion. Limitations inherent with electromechanical motion control and encoding are compactness in a portable device, slip, and inertial effects. Slip and inertial effects limit the ability of electro-mechanical hand scanners to respond to natural hand motions, which can peak at 16 inches per second or more in a single freehand swipe starting and ending across an 8 1/2" by 11" document.

An electro-optical solution overcomes the limitations of electromechanical motion control and encoding to permit image capture under unconstrained planar motion of translation and rotation as a user swipes the device across a document. Suitable on-board processing capability allows the device to track motion under all reasonable conditions of speed and directional change and reconstruct an undistorted image.



Figure 1. HP CapShare 920 Information Appliance

Introduction

In the past, typical hand scanners have not achieved a high level of usability because they constrain the user to a series of carefully-controlled linear scans and depend upon a PC for page reconstruction and image storage. In these devices,

Basics of Freehand Capture

Freehand image capture involves sweeping a linear array of pixel sensors over an image under full three degrees-of-freedom planar motion while keeping track of the rectilinear displacements of the end-points of the array each time a line of pixels is acquired. The path the image sensor takes is called a *swath*, and a complete document can be captured with a small sensor by passing it over the document in multiple, continuous swaths with a small amount of overlap without leaving gaps. This process produces a series of position-tagged rows of pixels which, if simply arranged in a stack, produce an image severely distorted when nonrectilinear motion is convolved with image capture as shown in Figure 2a. Here, the sensor is moved down the left side of a document, across the bottom, and back up along the right side with overlap. In principle, the path of the endpoints of the array over the document can be reconstructed from the position tags, and the pixel data can be deconvolved from planar motion. This produces an undistorted image where each pixel is placed in its correct location in a rectangular image array aligned with document coordinates. Figure 2b shows the reconstructed, stitched image from the raw data shown in Figure 2a.



Figure 2a. Stacked Rows of Pixel Data during a Freehand Scan



Figure 2b. Reconstructed Image

A device remaining in contact with the document as it is swiped across the surface has several advantages over "hovering" solutions, such as a handheld digital camera. A contact image sensor ("CIS") provides control over the imaging process by eliminating three of the six degrees-of-freedom that confound camera-based document capture. A CIS provides uniform illumination, constant focal distance and magnification, and assures alignment between the optical and page-normal axes. A linear sensor may provide high image resolution without the cost and complexity of an area sensor.

In the HP CapShare 920 information appliance, a linear CIS with photodetectors, optics, and LED illumination capture successive lines of pixels. To conserve battery power, the LEDs are flashed as any portion of the CIS moves to the next pixel position. To capture an image at 16 inches per second, gray-level pixel data rates of 18Mbits/sec are required.

The enabling element for freehand image capture is an optoelectronic motion encoder, or navigator. It correlates successive images of microscopic media surface features¹ to measure incremental motion. Accuracy of 0.8 um can be achieved at 16 inches per second. Grazing IR illumination enhances the contrast of surface features, and this useful for documents with dark backgrounds and areas covered by fused electrophotographic toner. The navigator consists of IR LEDs, imaging optics, and an ASIC with an imaging array, ADC, and digital processor and memory. Media surface features on the scale of 100 um are projected onto the array, and correlation between this comparison image and a reference image determines the offset between images. The process involves 200 million pixel correlations per second. The reference image is refreshed as the field of view changes with motion of the navigator across the surface.

The layout of the HP CapShare 920 imaging head is shown in Figure 3. While the navigators are not located at the ends of the CIS, a simple coordinate transformation gives their position during image capture.

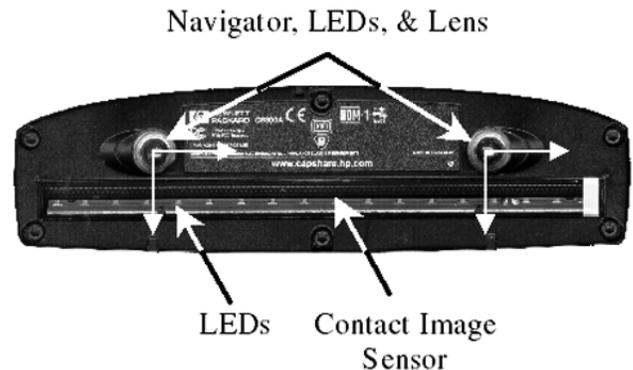


Figure 3. HP CapShare 920 Imaging Head

¹ Texture on this scale is found on common office papers, ink jet, laser printer, and copier papers, forms, books, newspapers, and magazines.

Navigation Algorithms

A sample digital image of illuminated paper fibers directly under the navigator is shown in Figure 4. A first image is used as the reference frame; subsequent images are the "comparison" frame. The navigator performs real-time correlation processing between these images using a nonzero-mean difference-detection function with the form

$$C_{i,j} = \sum_{m,n} |r_{m,n} - c_{m-i,n-j}|^2 \quad (1)$$

where $r_{m,n}$ ($c_{m,n}$) is the digitized value of the reference (comparison) image at pixel $\{m, n\}$, and i, j represent the shift of the comparison frame relative to the reference frame, measured in units of pixels.

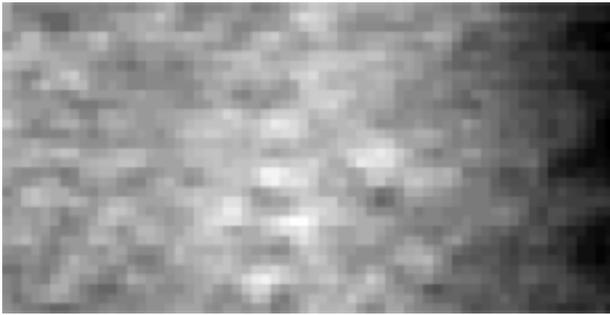


Figure 4. Navigator Image of White LaserJet paper

Application of this correlation function to a collection of relatively displaced images of the same geographic region of paper fibers yields a "bowl-shaped" correlation surface like that shown in Figure 5. Any successful navigation scheme must first build an accurate numerical model of this correlation surface. Although the complexity of the model must be constrained by requirements of the real-time navigation system, the scope of the model itself limits the ultimate accuracy with which the navigator can estimate the relative position of the sensor. With this compromise in mind, we attempt to represent the correlation bowl with an elliptic paraboloid; this quadratic model directs our intuition of the behavior of the navigation algorithm, and then allows us to place the navigation problem on a firm mathematical foundation.

The choice of a second-order function of x and y as an initial *idealized* three-dimensional model representation of the navigator correlation surface is motivated by considering a simple elliptic paraboloid described by

$$z' = f'(x', y') = q_{20} x'^2 + q_{02} y'^2. \quad (2)$$

If $q_{20}, q_{02} > 0$, then the surface has a unique minimum $z' = 0$ at $\{x', y'\} = \{0, 0\}$. In this case, the major and minor axes of the ellipse formed by the intersection of $f'(x', y')$ with a particular contour plane $z' = z_0$ are given by $(z_0/q_{20})^{1/2}$ and $(z_0/q_{02})^{1/2}$. Unfortunately, a correlation surface produced by actual paper fiber data is not so simple, and distortions must be taken into account. This can be done well enough by a general two-dimensional affine transformation and a subsequent three-dimensional translation mapping a real

correlation surface in $\{x, y, z\}$ into $\{x', y', z'\}$. By inverting both the translation and the affine transformation (in order), we can express the original (primed) coordinates of the unperturbed correlation surface in terms of $\{x, y, z\}$ as

$$x' = b_{11}(x - x_0) + b_{12}(y - y_0) \quad (3a)$$

$$y' = b_{21}(x - x_0) + b_{22}(y - y_0) \quad (3b)$$

$$z' = z - z_0 \quad (3c)$$

where $\{x_0, y_0, z_0\}$ are the components of the 3D translation vector, and $\{b_{11}, b_{12}, b_{21}, b_{22}\}$ are the elements of the matrix representing the 2D affine transformation.

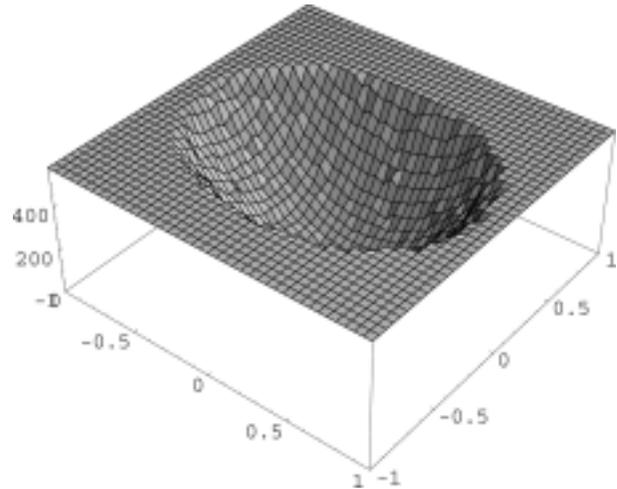


Figure 5. Result of applying Eq. (1) to data of Fig. 4

To produce a functional form for the transformed model surface in the $\{x, y, z\}$ coordinate system, substitute Eqs. (3a)–(3c) into (2) to obtain $z = f(x, y)$, where

$$f(x, y) = a_{00} + a_{10}x + a_{01}y + a_{20}x^2 + a_{11}xy + a_{02}y^2, \quad (4)$$

where

$$a_{20} = q_{20} b_{11}^2 + q_{02} b_{21}^2, \quad (5a)$$

$$a_{02} = q_{20} b_{12}^2 + q_{02} b_{22}^2, \quad (5b)$$

$$a_{11} = 2(q_{20} b_{11} b_{12} + q_{02} b_{21} b_{22}), \quad (5c)$$

$$a_{10} = -(a_{11} y_0 + 2 a_{20} x_0), \quad (5d)$$

$$a_{01} = -(a_{11} x_0 + 2 a_{02} y_0), \text{ and } \quad (5e)$$

$$a_{00} = z_0 + a_{11} x_0 y_0 + a_{20} x_0^2 + a_{02} y_0^2. \quad (5f)$$

Note that the transformed surface in the $\{x, y, z\}$ coordinate system has six terms of the form $a_{mn} x^m y^n$, where $\{m, n\} \in \{0, 1, 2\}$ and $0 \leq m + n \leq 2$. Furthermore, the two coefficients a_{10} and a_{01} can be uniquely expressed in terms of the displacements $\{x_0, y_0\}$, allowing the inversion of Eqs. (5d) and (5e) to find

$$x_0 = (a_{01} a_{11} - 2 a_{10} a_{02}) / (4 a_{20} a_{02} - a_{11}^2), \quad \text{and (6a)}$$

$$y_0 = (a_{10} a_{11} - 2 a_{01} a_{20}) / (4 a_{20} a_{02} - a_{11}^2). \quad (6b)$$

In other words, if the five coefficients a_{10} , a_{01} , a_{20} , a_{11} , and a_{02} can be determined numerically, then the components of the translation vector $\{x_0, y_0\}$ can be computed without explicitly finding either the elements of the inverse transformation matrix b or the two coefficients q_{20} and q_{02} .

Although the choice of the second-order form of $f(x, y)$ was motivated by considering the effects of affine transformations on an elliptic paraboloid, our approach can be generalized to apply to any correlation function in the neighborhood of a minimum (or maximum). Consider an arbitrary two-dimensional function which has the value $g(x_0, y_0)$ at the point $\mathbf{r}_0 = \{x_0, y_0\}$. In the neighborhood of that point, the effect of the small displacement vector $\mathbf{r} = \{x, y\}^T$ on the value $g(\mathbf{r}_0)$ is formally given by the Taylor series expansion[1]

$$\begin{aligned} g(\mathbf{r}_0 + \mathbf{r}) &= g(\mathbf{r}_0) + \nabla_0 g(\mathbf{r}_0) \cdot \mathbf{r} + \frac{1}{2} \mathbf{r}^T H(\mathbf{r}_0) \mathbf{r} + \dots \\ &= a_{00} + a_{10} x + a_{01} y + a_{20} x^2 + a_{11} x y + a_{02} y^2, \end{aligned} \quad (7)$$

where $H(\mathbf{r}_0)$ is the Hessian matrix evaluated at \mathbf{r}_0 , and now $a_{nm} = \partial^{m+n} g(\mathbf{r}_0) / \partial x_0^m \partial y_0^n$.

Hence, the choice of the second-order model function given by Eq. (4) is equivalent to the assumption that in the region of interest the "true" correlation surface can be accurately represented by a Taylor series expansion. If this assumption is significantly violated by a large displacement, then the translation vector computed by the algorithm above will *not* be reliable because at large distances the true surface will begin to diverge significantly from the second-order approximation given by Eq. (7).

Figure 6 shows the absolute residual errors obtained when the coordinates $\{x_0, y_0\}$ are calculated using Eqs. (6a) and (6b), after a numerical fit to correlation data computed from precisely displaced paper fiber images similar to Fig. 4.² The arrows point from the known position of the navigator measured using a high-precision mechanical stage to the final position computed by the embedded navigation algorithm. The RMS error is 0.02 pixel (0.8 μm), and the maximum error — found in the small regions near the corners where the small-displacement assumption begins to fail — is 0.07 pixel (3 μm).

Conclusion

The HP CapShare technology has fundamentally redefined the concept of portable, casual capture with a handheld device. With an IRDA port, the HP CapShare 920 can capture text and image documents and send them anywhere for subsequent viewing or printing. Novel hardware and software technologies have been developed that enable subpixel navigation for highly accurate real-time document

reconstruction of data swaths captured with convenient freehand motions.

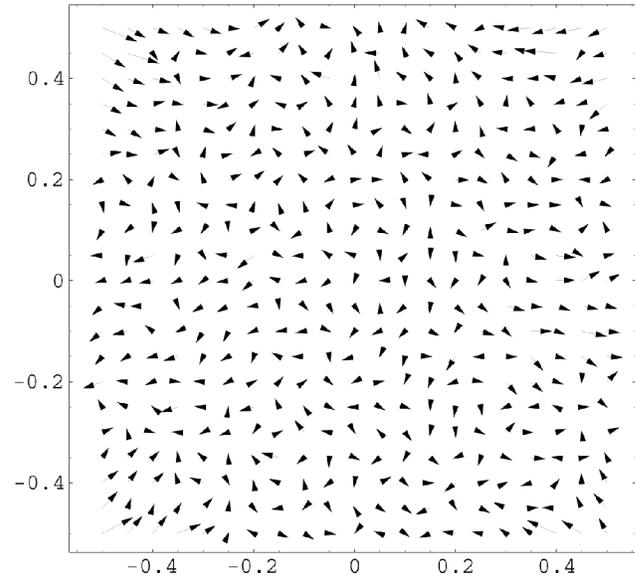


Figure 6. Absolute Residual Navigation Errors

References

1. Tom M. Apostol, **Calculus**, Vol. II, Second Edition (John Wiley & Sons, New York, 1969), pg. 308.

Biography

Dr. Ross R. Allen earned BS, MS, and Ph.D. degrees in mechanical engineering from the University of California, Davis. He joined Hewlett-Packard in 1981 to work on HP's thermal ink jet technology, and managed several projects on advanced TIJ systems. He transferred to HP Labs in 1989 where he invented and managed the development of the image capture technology used in the HP CapShare 920. He is now a Department Manager responsible for research and development in digital photography, hardcopy sensors, and document finishing.

Dr. Raymond G. Beausoleil earned the BS degree in Physics from the California Institute of Technology, and the MS and Ph.D. degrees in Physics from Stanford University. He conducted research and development in the fields of laser physics and nonlinear optics prior to joining Hewlett-Packard Laboratories in 1996. He invented the navigation and path reconstruction algorithms used in the HP CapShare 920. He is currently responsible for research and development in computational hardcopy, with applications to grayscale and color document imaging and reproduction and digital photography.

² A final post-process correction precalculated for white LaserJet paper has also been applied.