

of possible intended user actions. List **905** includes, for each action, a description of the parameters of an input corresponding to the user action. Processor **404** applies recognition function **903** $r_u(a, v)$ for each user action u in list **905**, and compares **904** the result to determine whether action u is deemed to have occurred.

[**0074**] For example, the visual feature vector v may include the height of the user's finger above the typing surface in, say, the five frames before the reference time stamp, and in the three frames thereafter, to form an eight-dimensional vector $v=(v_1, K, v_8)$. Recognition function **903** could then compute estimates of finger velocity before and after posited landfall by averaging the finger heights in these frames. Vision postulates the occurrence of a finger tap if the downward velocity before the reference time stamp is greater than a predefined threshold, and the velocity after the reference time stamp is smaller than a different predefined threshold. Similarly, the vector of acoustic features could be determined to support the occurrence of a finger tap if the intensity of the sound at the reference time stamp is greater than a predefined threshold. Mechanisms for determining this threshold are described in more detail below.

[**0075**] Signal **906** representing the particulars (or absence) of a user action, is transmitted to PDA **106** as an input to be interpreted as would any other input signal. One skilled in the art will recognize that the description of function **903** $r_u(a, v)$ is merely exemplary. A software component may effectively perform the role of this function without being explicitly encapsulated in a separate routine.

[**0076**] In addition, processor **404** determines features of the user action that combine parameters that pertain to sound and images. For instance, processor **404** may use images to determine the speed of descent of a finger onto surface **50**, and at the same time measure the energy of the sound produced by the impact, in order to determine that a quick, firm tap has been executed.

[**0077**] The present invention is capable of recognizing many different types of gestures, and of detecting and distinguishing among such gestures based on coincidence of visual and auditory stimuli. Detection mechanisms for different gestures may employ different recognition functions $r_u(a, v)$. Additional embodiments for recognition function **903** $r_u(a, v)$ and for different application scenarios are described in more detail below, in connection with **FIG. 3**.

[**0078**] Virtual Keyboard Implementation

[**0079**] The present invention may operate in conjunction with a virtual keyboard that is implemented according to known techniques or according to techniques set forth in the above-referenced related patents and application. As described above, such a virtual keyboard detects the location and approximate time of contact of the fingers with the typing surface, and informs a PDA or other device as to which key the user intended to press.

[**0080**] The present invention may be implemented, for example, as a sound-based detection system that is used in conjunction with a visual detection system. Referring now to **FIG. 1**, acoustic sensor **402** includes transducer **103** (e.g., a microphone). In one embodiment, acoustic sensor **402** includes a threshold comparator, using conventional analog techniques that are well known in the art. In an alternative embodiment, acoustic sensor **402** includes a digital signal

processing unit such as a small microprocessor, to allow more complex comparisons to be performed. In one embodiment, transducer **103** is implemented for example as a membrane or piezoelectric element. Transducer **103** is intimately coupled with surface **50** on which the user is typing, so as to better pick up acoustic signals resulting from the typing.

[**0081**] Optical sensor **401** generates signals representing visual detection of user action, and provides such signals to processor **404** via synchronizer **403**. Processor **404** interprets signals from optical sensor **401** and thereby determines which keys the user intended to strike, according to techniques described in related application "Method and Apparatus for Entering Data Using a Virtual Input Device," referenced above. Processor **404** combines interpreted signals from sensors **401** and **402** to improve the reliability and accuracy of detected keystrokes, as described in more detail below. In one embodiment, the method steps of the present invention are performed by processor **404**.

[**0082**] The components of the present invention are connected to or embedded in PDA **106** or some other device, to which the input collected by the present invention are supplied. Sensors **401** and **402** may be implemented as separate devices or components, or alternatively may be implemented within a single component. Flash memory **105**, or some other storage device, may be provided for storing calibration information and for use as a buffer when needed. In one embodiment, flash memory **105** can be implemented using a portion of existing memory of PDA **106** or other device.

[**0083**] Referring now to **FIG. 2**, there is shown an example of a physical embodiment of the present invention, wherein microphone transducer **103** is located at the bottom of attachment **201** (such as a docking station or cradle) of a PDA **106**. Alternatively, transducer **103** can be located at the bottom of PDA **106** itself, in which case attachment **201** may be omitted. **FIG. 2** depicts a three-dimensional sensor system **10** comprising a camera **506** focused essentially edge-on towards the fingers **30** of a user's hands **40**, as the fingers type on typing surface **50**, shown here atop a desk or other work surface **60**. In this example, typing surface **50** bears a printed or projected template **70** comprising lines or indicia representing a keyboard. As such, template **70** may have printed images of keyboard keys, as shown, but it is understood the keys are electronically passive, and are merely representations of real keys. Typing surface **50** is defined as lying in a Z-X plane in which various points along the X-axis relate to left-to-right column locations of keys, various points along the Z-axis relate to front-to-back row positions of keys, and Y-axis positions relate to vertical distances above the Z-X plane. It is understood that (X,Y,Z) locations are a continuum of vector positional points, and that various axis positions are definable in substantially more than the few number of points indicated in **FIG. 2**.

[**0084**] If desired, template **70** may simply contain row lines and column lines demarking where keys would be present. Typing surface **50** with template **70** printed or otherwise appearing thereon is a virtual input device that in the example shown emulates a keyboard. It is understood that the arrangement of keys need not be in a rectangular matrix as shown for ease of illustration in **FIG. 2**, but may be laid out in staggered or offset positions as in a conven-