



(22) Date de dépôt/Filing Date: 2002/03/05

(41) Mise à la disp. pub./Open to Public Insp.: 2002/09/05

(45) Date de délivrance/Issue Date: 2008/05/20

(30) Priorité/Priority: 2001/03/05 (US09/799,299)

(51) Cl.Int./Int.Cl. *G09B 7/00* (2006.01),
G06F 17/00 (2006.01)

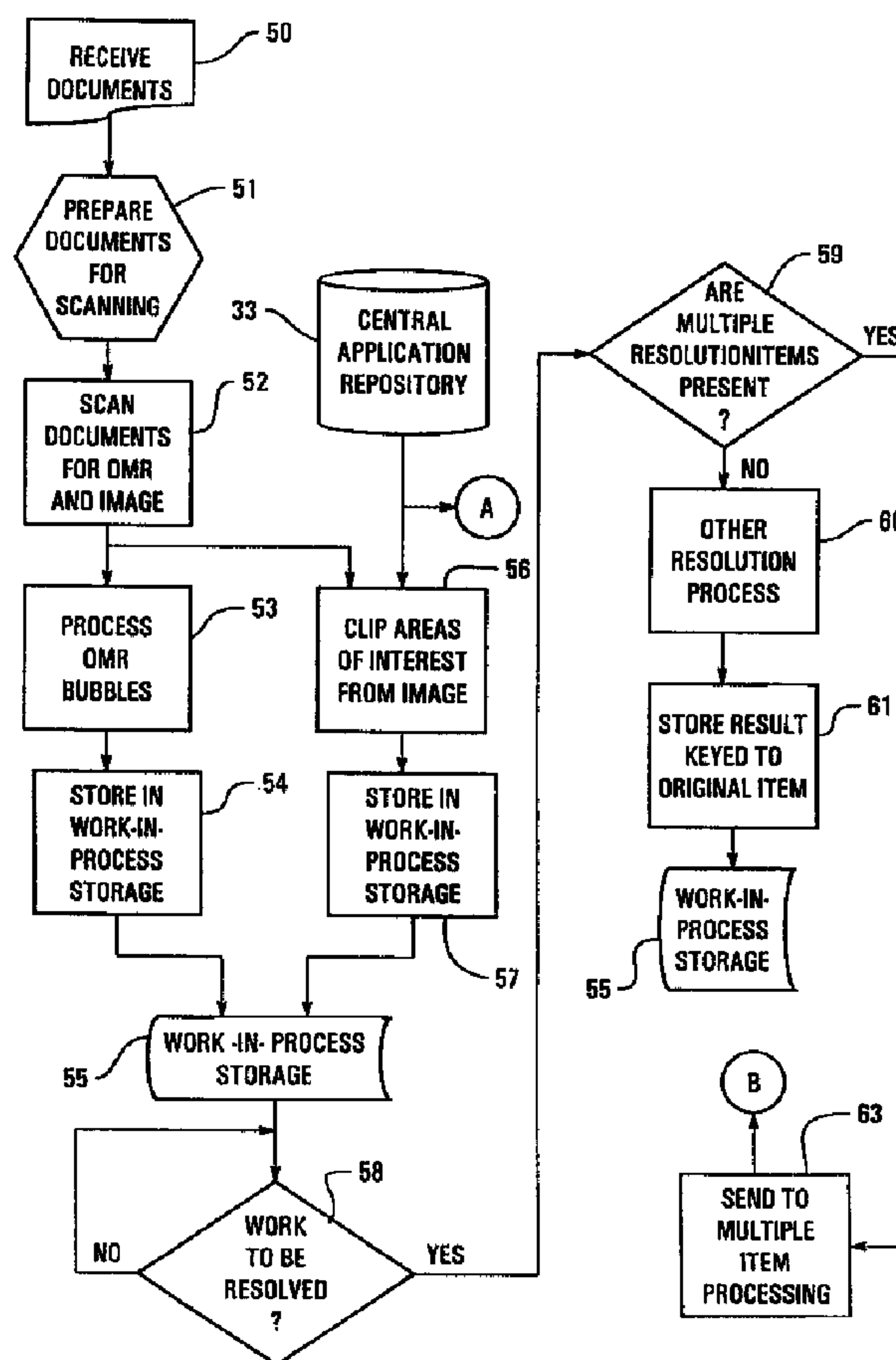
(72) Inventeurs/Inventors:
KNOWLES, KRISTIAN, GB;
PASCALE, ROBERT, AU

(73) Propriétaire/Owner:
NCS PEARSON, INC., US

(74) Agent: BLAKE, CASSELS & GRAYDON LLP

(54) Titre : SYSTEME DE CONTROLE DES REPONSES AUX QUESTIONS D'EXAMEN

(54) Title: TEST QUESTION RESPONSE VERIFICATION SYSTEM



(57) Abrégé/Abstract:

The present invention relates to a system and method for verifying and editing optically imaged responses received from a response provider.



Abstract of the Invention

The present invention relates to a system and method for verifying and editing optically imaged responses received from a response provider.

TEST QUESTION RESPONSE VERIFICATION SYSTEM

Field of the Invention

5 The present invention relates to a system and method for verifying and editing optically imaged responses received from a response provider.

Background of the Invention

10 The scoring of test question responses that cannot be scored by machine is typically carried out manually or by presentation on a computer monitor. Manual scoring involves a human manually scoring a physical test question response sheet. Scoring by presentation to a human of the test question responses using a computer involves scoring an electronic representation of test question response or responses presented to a scorer via a computer monitor or other machine that can be programmed to manipulate symbols. It is the latter scoring procedure to which the present invention
15 relates.

 In order to present a test question response to a scorer viewing a computer monitor, several preparation steps typically occur to enable the scorer to receive the test question response, view it, score it and record the score with the necessary precision, speed and accuracy required in the test scoring industry. With test
20 processing that analyzes optically imaged test question responses, scanners, facsimile machines and other optical imaging devices known to those skilled in the art are used to create an electronic image of the test question response that was provided by a test taker. The electronic images may be broken down into smaller images generally referred to as image clips. The electronic images and image clips may be stored in a
25 computer or other storage media known to those skilled in the art. The electronic images and image clips are typically extracted to data using well known and commercially available optical character recognition software, image character recognition software and other similar computer programs. The data can then be utilized in a number of ways to aid the test scoring process.

While character recognition software programs have improved greatly, a need still exists to verify the results of the image conversion. Verification is especially important in the test scoring industry where test takers' scores may depend on the accuracy of the data extracted from the electronic images of the test question responses.

5 Most character recognition software programs provide an automatic verification process performed electronically; however, even those results often are subject to a relatively high rate of error. In many situations, human editors use computer monitors at remotely-located geographic locations to verify the accuracy of the data extracted from the electronic images. A workflow system may be utilized to present to the editors data

10 or images, or both, from which they are to verify accuracy.

Several problems using human editors to verify converted images have been identified. Editors often lack particular skills desired to analyze all types of data that may have been extracted from images. For example, if the test taker used Chinese characters to indicate his/her test question responses, editors not fluent in Chinese

15 would find it difficult to analyze the information presented to them. Similarly, editors not comfortable with the numerous and complex symbols used on math tests may find it difficult to analyze the math test question responses presented to them.

The slow speed at which editors typically have analyzed the data extracted from images has also been a problem. Editors were randomly presented with

20 data and/or images in need of analysis. For instance, an editor might be first presented with the image or the image and data of the test taker's first name. Rather than presenting a last name or middle initial to the editor next, the editor might be presented with the image or the image and data of a math test question response. The random nature of the materials presented to editors slowed their process in verifying the data

25 extracted.

Summary of the Invention

The present distributed test processing workflow system facilitates accurate, consistent and high quality image capture, conversion to data and verification of test question responses provided by a test taker.

The present verification editing method divides test question response images into smaller images or "image clips" for conversion to data using well known image recognition software. In the method, it is determined whether the particular data obtained from the image clips requires verification. If verification is required,
 5 information related to each image clip is processed against particular appropriate skills of human editors in order to select human editors having particular appropriate skills. Selected image clips are then recompiled into coherent image portions of the test question response and selectively presented, with or without corresponding data, to the human editor for analysis. Finally, the decisions of the human editor are electronically
 10 recorded.

Brief Description of the Drawings

FIG. 1 is a hardware diagram of a distributed workflow system as possible using the claimed inventions.

FIG. 1A is a three tier architecture view exemplary of the type used in
 15 the claimed inventions.

FIG. 2 is a block diagram of a portion of the network shown in FIG. 1.

FIG. 3 is a block diagram of the scanning configuration in the network of
 FIG. 1.

FIG. 4 is a block diagram of a portion of the network of FIG. 1.

20 FIG. 5 is a flow chart of receiving and processing of response items.

FIG. 6 is a flow chart of one embodiment of the workflow tracking invention.

FIG. 7 is a flow chart of one embodiment of the pre-data-collection applications invention:

25 FIGs. 8, 8A, 8B is a flow chart of one embodiment of the verification invention.

FIG. 9 is a flow chart of one embodiment of the multiple server response processing invention.

30 FIG. 9A is a block diagram of one embodiment of the multiple server response processing invention.

FIG. 10 is a flow chart of one embodiment of the image archive invention.

Detailed Description of the Preferred Embodiments

General Overview

5 The present invention pertains to a response document processing workflow system. Response document processing is often employed in the test grading and survey analysis industries. By way of explanation, but not limitation, a need has developed in the response document processing industry to process more documents, faster, with better accuracy and less expensively than ever before. Electronic imaging
10 has become very popular in the response document processing industry given the flexibility, speed and efficiency gained through the use of electronic images over paper documents. Scanners are one popular type of optical imaging device; another is the facsimile machine. The increased volume of documents to be processed, and the quicker turn-around time, requires the combination of multiple scanners, multiple
15 processing units and multiple operator stations working in parallel or in series to quickly finish a processing job. A network is often employed to couple the various hardware components. The networked response processing workflow system described below satisfies the above described needs.

 The beginning of a response processing project or "job" may occur well
20 before the documents containing the responses are received at a response processing center. For example, customer or respondent information such as name, mailing address, identification number, and other personalized information may be printed onto response forms, such as test forms or survey forms. The personalized forms may then
25 be provided to a customer or directly to a response provider with or without a personalized cover letter. The above is accomplished by populating and maintaining a database or other electronic storage system with personalized information necessary for personalization of the response forms.

 Whether the response documents received at the response processing center are those that were personalized or not, it is critical that all documents received
30 are processed and accounted for. An audit trail is therefore started at document receipt time. The processing center will collect the documents in pre-selected quantities or

batches. A batch header document containing a unique bar-code batch number or other similar machine readable identifier will be assigned to each batch of documents.

The documents will be optically imaged at the processing center in one batch or subdivided into smaller batches and imaged sequentially. Multiple processing centers may be used to meet customer needs. While scanning is the preferred method of optical imaging, facsimile and other optical imaging devices known to those skilled in the art may be used to create an electronic image. After imaging, the physical documents typically are stored in a warehouse, their location identified through the unique bar-code batch number or other machine readable identifier on the batch header.

Several steps may be required prior to collection of data from the electronic image taken from each document. For example, image enhancement and image identification are available applications that may be implemented prior to data collection. Data collection or recognition is accomplished using one or more recognition technologies, for which an interface is provided in the system. The system must be flexible enough to handle responses in many different languages and disciplines. Because accuracy is critical in testing and response imaging, human verification may be applied to any data derived from electronic images that is invalid or uncertain. The customer may require multiple verification. In such a case, if the system detects a discrepancy between the values of the verified data, the image and the data values are automatically sent to a third adjudicator for resolution (adjudication). In an alternative embodiment, a double adjudication embodiment, the data values are automatically sent to a third and fourth adjudicator for resolution. If the adjudicators disagree, the image is sent to a manager or other NCS official for final determination.

After the data has been verified, it may be sent to other systems for processing consistent with tests, such as grading; surveys, such as tallying; or other processes known to those skilled in the art.

Although electronic storage capabilities have helped lower operating costs, the need remains to minimize the cost of storing electronic images. An image archive component of the workflow system is provided. It can be appreciated that the archives will play an important role in the event that a dispute arises over the accuracy

of data from an imaged response or test. Further, imaged archives can be accessed by customers wanting to view the original scanned document for any number of reasons.

Hardware Overview

FIG. 1 illustrates an example of a hardware configuration for the present response document processing system. This configuration is shown as an example only; many different hardware configurations are available for implementing the response processing functions described below, and others will become available as hardware technology advances, as recognized by those skilled in the art. The network shown is generally described as a three tier or three layer architecture, (See FIG 1A for more detail on three tier architecture) including a bottom tier or data layer of at least one server computer. The data layer may include a server subsystem including several servers as shown in Fig. 1 at 10-13. Separate servers may be used for such tasks as maintaining a report database 10, a data capture database 11, a local area network server 12, archived images database 13 and a "jukebox" 14 - a hardware mechanism for allowing access to one of a group of discs, especially CD-ROMs or other storage media. The data layer of the hardware system is interfaced with a middle tier or business objects layer.

The business objects layer provides process management where business logic and rules are executed and can accommodate hundreds of users (as compared to only 100 users with a two tier architecture - also usable in the present system) by providing functions such as queuing, application execution, and database staging. In Fig. 1 the business objects layer includes a site server 15 interfaced through a hub computer and a site server 16 to application servers 17 and 18. The application servers in the preferred embodiment utilize Microsoft's Transactional Server software. The business objects layer interfaces through hub computer 20 to a presentation layer or top tier. The presentation layer of the preferred embodiment includes computers 21 interfaced to the system via a local area network or LAN, a computer 22 and a router 23 interfaced to the system via a wide area network or WAN and a computer 24 interfaced to the system via the Internet or other global computer network. The internet connection also includes a firewall 25 and router 26. The optical imaging devices 27 (a scanner and facsimile machine) are shown in this embodiment coupled to client

computers 21. In an alternative embodiment, the optical imaging devices shown, a scanner or facsimile machine 27 may be coupled with computers 22 or computer 24.

The system uses the optical imaging devices 27, usually scanners, in the presentation layer to optically image responses provided by respondents on questionnaires, surveys, tests or other similar response documents on which responses are received. These response documents may comprise, for example, test forms with "bubbles" or ovals representing possible answers to test questions, handwritten essays, survey responses, or other various types of written or printed or marked information provided by the respondent. After receiving the optically imaged responses, the system can prepare those electronic images for subsequent processing, such as test grading or survey analysis. A response item is, therefore, an electronic representation of at least a portion of a questionnaire, legal document or response document, including, but not limited to, a test form with answers, a contract or a survey. The system and methods described below are then employed to process the response items such that they are prepared for subsequent online analysis. See, for example, U.S. Patents 5,752,836; 5,735,694; 5,718,591; 5,716,213; 5,709,551; 5,690,497; 5,321,611; 5,458,493; 5,433,615; 5,437,554; 5,466,159; 5,452,379; 5,672,060; 5,987,149; and 5,558,521 describing online training, scoring, teaching and reporting systems and methods. See also, U.S. Patents 5,711,673; 5,420,407; 5,262,943; 5,262,624; 5,184,003; 5,134,669; 5,086,215; 4,934,684; 4,857,715; and 4,217,487.

The system must be capable of supporting various types of optical imaging devices, such as intelligent scanners like the NCS 5000i™ and non-intelligent scanners that utilize the Kofax/ISIS/TWAIN™ interface. The client subsystem computers 21, 22 and 24 typically initially receive the response items and may distribute them to the business objects layer server subsystem before sending them to the data layer server subsystem for storage, archival or other processing. For example, a response document optically imaged by the scanner 27 will be assigned a unique file name and digital identifier at the client subsystem computers 21, 22 or 24, then sent via the business objects layer server subsystem to the data layer server subsystem for storage and further

processing using the system and methods described below. It is also possible for servers 10-13 to transfer the response items via a global, local or other network known to those skilled in the art to another server subsystem in a remote location. The ability to share the response items with another server subsystem makes work sharing and efficient use of resources available and practical. The servers may be in the same physical area as one another or separated by great distances, such as between different countries or continents. For example, the advantages of different time zones can be utilized using multiple servers and work sharing so as to avoid down time and delay in processing the customer's project.

FIG. 2 is a more detailed block diagram of a portion of the network shown in FIG. 1. As shown in FIG. 2, the optical imaging device 27 shown in FIG. 1, is typically a scanner 27 interfaced to a computer 28 or client subsystem computers 21, 22 or 24. The scanner 27 contains a camera 29 for reading through optically imaging all or portions of a response document. For example, using "clipping" to read individual responses on a response documents or reading all the responses including all or some of the document background of the response document (full page imaging).

FIG. 3 shows a more detailed block diagram of a typical scanning unit. Some scanners have camera optics and filter configurations that are optically "blind" to certain colors, typically called drop-out colors. As the scanner cannot see these colors and normally these colors constitute the response form background or template that the scanner will not capture the background of the form within the electronic image.

One or more client computers 21, 22 or 24, preferably implemented with the most current high performance computer chip and memory, usually contain a frame buffer 251 for receiving the scanned image data from the scanner 27. The computer 28, which is optional, is typically interfaced to the scanner 27 and client computers 21, 22 or 24 for controlling the operation of the optical imaging device, in this case the scanner 27. The system may alternatively be configured such that all of the functionality of the computer 28 is with client computers 21, 22 or 24. The computer 28, if used, controls the scanner and thus controls when image data is scanned in and subsequently transferred to client computers 21, 22 or 24. The client computers essentially act as a buffer for temporarily holding the image data. The client computers can also "clip"

areas of interest from the electronic image. Clipping involves electronically removing, typically in software, a portion of the response item or scanned image. These clipped areas may comprise any portion of a response item: e.g., a handwritten essay or selected response positions. For example, if scanner 27 is an intelligent scanner such as the NCS 5000i, "clipping" of an area of interest from the image during scan time and representing at least a portion of the scanned response is possible. Examples of two systems for storing and extracting information from scanned images of test answer sheets are shown in U.S. Pat. Nos. 5,134,669 and 5,103,490, both of which are assigned to National Computer Systems, Inc.

10

FIG. 4 is a block diagram of the hardware and software functions in a server in the network of FIG. 1. A scan control module 31 interfaces with client computers 21, 22 or 24 and receives the image data. The image data is stored in a raw item database 36. For example, a SQL compliant industry standard relational database such as Microsoft SQL Server, may be used. The system typically further utilizes open systems architecture such as ODBC, to provide the database access mechanism. In an alternative embodiment, an entry level database is utilized thus removing the cost of a database such as Microsoft SQL Server. The central application repository (CAR) 33 typically stores document definitions and handling criteria. The document process queue 37 function as a buffer into a main processing module 45 preferably in servers 10-13. The main processing module 45 controls the processing of response items. It controls the transmission of response items to client computers 21, 22 and 24.

15

20

Software Overview

Although many programming languages and classes may be utilized to carry out the present inventions, one of the preferred embodiments uses object oriented programming. Object oriented programming includes writing programs in one of a class of programming languages and techniques based on the concept of an "object" which is a data structure encapsulated with a set of routines, called "methods" which operate on the data. (See FIG. 1A). Operations on the data can only be performed via these methods, which are common to all objects that are instances of a particular "class." Thus the interface to objects is well defined, and allows the code implementing

25

30

the methods to be changed so long as the interface remains the same. For example, an Unknown interface is shown in FIG. 1A. Each class is a separate module and has a position in a class hierarchy. Methods or code in one class can be passed down the hierarchy to a subclass or inherited from a superclass. Procedure calls are described in terms of message passing. A message names a method and may optionally include other arguments. When a message is sent to an object, the method is looked up in the object's class to find out how to perform that operation on the given object. If the method is not defined for the object's class, it is looked for in its superclass and so on up the class hierarchy until it is found or there is no higher superclass. Procedure calls always return a result object, which may be an error, as in the case where no superclass defines the requested method. FIG. 1A represents one overview of the object oriented programming structure of the present invention. Others object oriented programming structures are contemplated and utilized.

FIG. 5 is a flowchart of a general overview of typical scanning and processing of response forms and responses. The document processing center receives the response sheets or other documents, at step 50 and performs initial clerical preparation of the documents at step 51 prior to scanning. For example, if the documents received include a digital identifier in the form of a bar code or other machine readable format, the digital identifier is read into the system using a barcode reader or other similar technology known to one skilled in the art. In an alternative embodiment, the documents received do not include a digital identifier. In the latter scenario, the receiver of the documents may manually organize the received documents into predefined groups or batches. Such batches are a collection of documents, or the electronic images and/or data derived from the documents, organized in some relational manner. For example, a university customer may request that all response sheets from freshman test takers comprise a batch separate from the response sheets from sophomores. Batches may include sub-batches. For example, the freshman batch may be further divided by the sex of the student/respondent.

The system at step 52 scans the documents comprising one batch or a sub-set of one batch. The system is designed to accommodate various types of optical imaging devices, as disclosed above. If the scanner is capable, the system may process

the OMR bubbles at step 53, during scan time, and store the data derived from
 “reading” the OMR marks in the work-in-process storage (WIP) at step 54. The system
 at step 56 can “clip” areas of interest from the document, again during scan time, saving
 data derived from the image in the WIP 54. The clipped image is saved as an image file
 5 in server 10 or client computer 12. The step of “clipping” involves electronically
 copying a portion of the scanned image. This can be done at scan time using an
 intelligent scanner or later if a non-intelligent scanner is used. These “clipped” areas
 may comprise any portion of a response sheet; for example a handwritten essay,
 machine printed name or graphical response such as a mark or drawing. The system
 10 may also receive image data directly from foreign sources, magnetic or electronic, and
 store the data in raw item database 36. Subsequent operations on the data are the same
 regardless as to the source of the data. After “clipping” areas of interest from the
 image, the system stores the response items at step 57 in the WIP 54 or transfers the
 response items to the server 10 for storage.

15 The system can run on any computer using a standard operating system
 such as Microsoft Windows™ or NT. The server subsystem typically operates Microsoft
 Transaction Server (MTS™) and Microsoft Internet Information Server (IIS 4.0™)
 connecting through ActiveX Data Access Object (ADO™) to connect to either MS SQL
 Server or MS Access™ (on NT Server) or Oracle™ (on NT or UNIX™) for the database.

20 It has been found that if the scanner 27 coupled to the client subsystem is
 an intelligent scanner such as the NCS 5000I, the client computer PC runs well using
 Windows NT. For programming and support efficiency, a standard object oriented
 programming language such as C++, MFC or Visual Basic are recommended.
 However, one skilled in the art will appreciate that almost any programming language
 25 can be used to accomplish the present invention. It preferably is also part of the present
 software to utilize the Internet as well as private Intranets and the flexibility of browsers
 such as Netscape Navigator and Microsoft Internet Explorer.

30 Prior to providing the electronic images to human resolvers for online
 grading or analysis, steps 58-63, the system performs additional preparation processes,
 described in detail below, to ensure the identity of the image, the quality of the image to
 be presented, the relationship of the image to the original document from which it was

imaged, the accuracy of data derived from the document imaged and the archival storage of the images for future use or reference or both.

Response Processing Workflow Tracking System And Method

FIG. 6 is a flowchart of a preferred embodiment of the response item
5 workflow tracking and organization invention.

A human administrator, in steps 601, 602 and 603, inputs information into a system database via a user interface. In step 601, information about the optical imaging device(s) to be used in imaging the response documents is entered. Such information may include optical imaging device identifier and interface type.
10 Information pertaining to the document types to be imaged is entered in step 602. Document information is preferably obtained by capturing an electronic image of a blank response form and a combination of commercially available and bespoke software to define document attributes such as regions of interest. In step 603 information specific to the particular job or project to be undertaken, including, but not
15 limited to, a narrative about the customer, workflow rules, customer logic and machine readable identifier, is entered by the human administrator. The above steps can, but are not required to, occur before receipt of the response documents provided by respondents which occurs at 604.

The documents received, and to be imaged, are assembled by the human
20 administrator, step 605, into batches or groups of a suitable size for processing consistent with the customer's needs. In an alternative, batches may be defined by the customer and if of a non-suitable size, may be divided into smaller sub-batches to facilitate imaging. In step 606, it is determined whether the documents received at the processing center include a machine readable identifier ("MRI"). A MRI may include a
25 bar code or other similar machine readable number or code. If a MRI is included on the documents to be imaged, the administrator will input batch information into the system, step 607, excluding the number of documents in the batch. Batch information may include text including a batch name, an optional batch name provided by the customer and an optional narrative text string. Using a bar code "reader" or other similar device,
30 the MRIs will be read and entered into the system database at step 608. In step 609, the documents are counted using the machine readable identifier and the number of

documents to be imaged is stored in a system database. The system generates, at step 611, a unique bar-code batch identifier for each batch assembled and prints a batch header document that includes the unique bar-code batch identifier as well as the batch information described above.

5 If the documents do not include a machine readable identifier, step 610, a human administrator manually counts and inputs the number of documents in the batch to be imaged as well as the batch information described above. The system generates, at step 611, a unique bar-code batch identifier for each batch assembled and prints a batch header document that includes the unique bar-code batch identifier as well as the
10 batch information described above.

 In step 612, the documents are optically imaged using scanner, facsimile machine or other optical imaging device known to those skilled in the art. Scanners can be grouped in two general categories: intelligent and non-intelligent scanners. It is a significant aspect of the system and method described herein to be adaptable to a variety
15 of optical imaging devices, including scanners. If an intelligent scanner, such as the NCS 5000i, is used to electronically image the documents, the system will "clip" data or images, step 613, from certain regions of interest on the response document. Preferably, only the data or response image is captured, and the response document background is not captured. More than one "clip" may be taken from a document. In
20 step 613, one multiple image file containing all the image clips from one document is created. In one preferred embodiment, the images are in tag imaged file format (TIFF). A custom tag typically is utilized that contains an individual digital identifier (IDI) that uniquely identifies the image clip. The IDI contains the information about the image that the system requires for subsequent display in the correct context. Typical items
25 included in the IDI are position coordinates and form identification. The system then stores any data derived from the image clips in a database, step 614. Storing the IDI in the tag of the image, rather than in an associated database, permits faster retrieval of the data when the image needs to be displayed. It has the added benefit of avoiding any risk of the image becoming separated from the relevant data.

30 If a non-intelligent scanner is used, the scanner typically produces an image file for each page of the document; one image per page side, step 615.

Interpretation of the data typically is not performed by a non-intelligent scanner. In step 616, the image is stored in files, a separate directory is created for each batch of documents.

Pre-Data Collection Preparation Of Responses

5 FIG. 7 is a flowchart of the preferred embodiment of the pre-data collection system and method applied to electronic images of responses in the processing workflow system. The pre-data-collection applications are used to effect various tasks upon the electronic images desired for subsequent data collection. For example, an electronic image of a response may require "cleaning" via commercially
10 available software to improve the electronic image from which data will be captured. In step 701 a human administrator generates an image-device-type command informing the system of the type of optical imaging device to be used to create electronic images of the responses. Optical imaging devices include scanners, facsimile machines and other devices known to those skilled in the art that can convert an image into binary
15 data. This preferred embodiment utilizes an optical imaging device that outputs binary data in a tagged image file format (TIFF). Upon selecting an optical imaging device type in step 702, the system receives a command informing it of the optical image device type to be used to electronically capture the test question responses as electronic images. The word "command" is meant in the broadest possible interpretation to
20 include a character string that tells a program to perform a specific action or supply input information. Specifically, the word "command" includes menu items as commands.

 In step 703, the human administrator is asked to select which of a number of pre-data-collection software applications are to be applied to the electronic
25 images of test question responses. Pre-data-collection software includes any software application that may be applied to an electronic image. In the preferred embodiment, pre-data-collection software applications include software programs with the ability to recognize handwriting, software programs with the ability to read a barcode, software applications with the ability to identify an image as belonging to a particular set of
30 images, and software applications with the ability to remove from electronic images noise and superfluous lines without removing characters, as well as other software

applications know to those skilled in the art of digitizing an optical image into an electronic image represented as binary data. Upon selection of the same, the system is updated via a pre-data-collection command informing it of the pre-data-collection software applications to be applied.

5 In step 704, the system begins to receive electronic images of the test question responses via the optical imaging device. The output of the optical imaging device is binary data and preferably, binary data capable of storage in tagged image file format (TIFF). The electronic images of test question responses are received in the server subsystem at step 705. The system then determines which one or more of the
10 plurality of pre-data-collection software applications to apply to the received electronic images based upon the optical-imaging-device-type-command and the pre-data-collection software command, step 706. In this described preferred embodiment, after determining which of the plurality of pre-data-collection software applications to apply to the electronic images, the received electronic images are returned to the client
15 computer system at step 707. The selected pre-data-collection software applications are applied sequentially to the returned electronic images in step 708.

Typically, the output from applying one or more of the plurality of pre-data-collection software applications will be stored in the server subsystem, step 709. For example, an electronic image that has been cleaned of noise and superfluous lines
20 without removing actual characters will be saved as an electronic image separate from the electronic image to which the pre-data-collection software application was applied. Similarly, the data returned from reading a barcode or identifying an image as belonging to a particular set of images would be saved in the server subsystem.

Following application of the pre-data-collection software programs, in
25 step 710, one or more data collection software applications typically then are applied to the returned electronic images. Data collection software applications include intelligent character recognition, optical character recognition and other software applications that recognize printed or written characters by a computer. Such programs are well know to those skilled in the art of image processing and character recognition. The text file or
30 other similar data generated from the application of data collection software is stored in the server subsystem at step 711.

In an alternative embodiment, the application of the pre-data-collection software program occurs in the server subsystem rather than the client computer system. Such an embodiment saves the step of returning the received electronic images to the client computer system; however, any benefit is offset by the usage of the server
 5 subsystem resources to apply the pre-data-collection software applications in the server subsystem.

Test Question Response Verification System And Method

FIGS. 8 and 8A are flowcharts of one preferred embodiment of a test question response verification system and method. Typically electronic images of test
 10 question responses can be divided into image clips, where an image clip is an electronic image smaller than a full-page image. Data is then typically converted from a binary representation of the image clip using image recognition techniques such as intelligent character recognition, optical character recognition or other character recognition and image processing programs known to those skilled in the art. As recognized by those
 15 skilled in the art, character recognition involves complex image processing algorithms and rarely achieves 100 percent accuracy. Accordingly proofreading or verification is recommended.

The system at steps 801 and 802 determines if data derived from electronic images having undergone character recognition requires verification. At step
 20 803, the system has determined that verification is needed and contemporaneously determines how many times each item must be independently verified. Typically a human administrator pre-determines the number of independent verifications any data requires. The system can be programmed to apply one or more verification passes to particular items. Each verification pass is an independent verification of the accuracy of
 25 the data converted from an electronic image. If only one verification pass is to be carried out, the system retrieves the data to be verified from the server subsystem at step 804 and makes the data available in a work queue at step 809.

If more than one verification pass is required, in addition to retrieving the data to be verified and placing the same in a work queue at steps 804 and 809, the
 30 system creates an empty field for each verification pass greater than one at step 805. For example, if three verification passes are required of a particular item, the system

will create two empty fields for verification passes 2 and 3. The system then retrieves an image of the information at step 806 that was converted into data at step 804. The system sends the images to work queues, one queue per each verification pass at step 809. The system then holds the data and images in work queues at step 809 until an editor logs into the server subsystem.

Step 820 of the system provides to a human editor a list of options available from which to select using a graphical user interface. One of the selections available includes verification. The human editor logs on to the system at step 821. At step 822 the system determines if the login procedure has been complied with by the human editor. If not, at step 823, the human editor is logged out of the system. If the login is accepted, at step 824, the editor is presented with a choice of particular editing jobs from which to select. A job, for example, may be a particular set of tests or surveys or questionnaires provided to NCS for processing. The jobs presented to the human editor are limited to those matching information known with the particular human editor. Once the particular job is selected, the system processes information related to each image clip against information pertaining to the particular skills of the human editor. In step 826, the system determines whether the human editor will be verifying an item undergoing only one pass of verification at step 826. If the item is only undergoing one pass of verification, at step 827, the system presents to the human editor selected image clips recompiled into coherent image portions such that the human editor sees on his or her computer monitor the complete background of a test form but only the selected images clips corresponding to material necessary for verification, typically the selected image clip and the corresponding data.

If the human editor is verifying an item undergoing more than one verification pass, the system at step 828 presents an image clip and its background, but not any derived text. The human editor then records a decision based on his or her observation of the accuracy of the material presented in step 829.

In step 831 the data verified is submitted to a compare queue. The system determines the number of verification passes the data underwent at step 832. If the data was subject to only one verification pass, then there is nothing with which to compare it and the data is marked complete at step 839. If however the data was subject

to more than one verification pass, then all the decisions entered by the human editors in step 829 are compared in step 835. In step 836, the system determines if all the decisions entered by the human editors are the same, and if they are the same the data is marked complete at step 839. If however, the system determines that all the decisions entered by the human editors are not the same at step 836, the system routes all entered decisions to an adjudication process.

Multiple Server Processing Workflow System And Method

FIG. 9 is a flowchart that describes the multiple server test processing workflow system and method. At step 901 documents containing test question responses are optically imaged. The optically imaged test question responses are received by a server computer, at step 902, coupled to the client computer system that includes the optical imaging device. In an alternative embodiment, the server or servers that receive the optically imaged test question responses may not be the server computer coupled to the client computer that includes the optical imaging device. The receiving server computer may or may not store electronic images. In step 903 the system distributes from the receiving server computer to at least a second one of the plurality of server computers all or a portion of the received electronic images. The server computer or computers to which the received electronic images are distributed may or may not store the electronic images. In step 904 the electronic images are analyzed in one or more of the receiving server computers. It is further contemplated that the analysis step may occur not only in one of or all of the server computers but also may occur in one of or all of the client computers coupled to a particular server computer. The computer or computers which perform the analysis of the electronic images may return to the work flow system results for collation with other test question response analysis results, see step 905.

In an alternative embodiment, the receiving server computers are configured to analyze particular test response subject matter including particular human languages or particular fields of study such as math, science or subjects that may benefit from test processing personnel having particular skills related to the subject matters.

Archiving Electronic Images Of Responses

FIG. 10 is a flowchart of the preferred embodiment of the test workflow image archive invention. In step 1001, an optical image of the test form template is created and stored in the server subsystem. In step 1002, the optical imaging device "clips" selected portions of a test question response from a test taken by a test taker generating an electronic representations of the test question responses. Typically, the output of the optical imaging device is in tagged image file format (TIFF); however, other formats may be used such as JPEG, BMP, etc. In step 1003, the server subsystem coupled to the optical imaging device receives the electronic representation of the test question responses and stores the same. Typically, the electronic representations of the test question responses are saved on a magnetic medium such as a computer disk, magnetic tape or other storage medium known those skilled in the art. To minimize the use of expensive archival resources, the test form template and electronic representation of test question responses are transferred to an archival storage medium and stored separately, step 1004. The archival storage media includes magneto-optical disks, optical disks and other storage media know to those in the art as suitable for archiving electronic images. An electronic index is generated in step 1005, where the index links the test form template to the electronic representations of electronic images - now both stored on the archival medium. The index is typically a relational database that allows the definition of data structures, storage and retrieval operations and integrity constraints. In such a database the data and relations between them are organized in tables. A table is a collection of records and each record in a table contains the same fields. Certain fields may be designated as keys, which means that searches for specific values of that field will use indexing to speed them up. The fields typically correspond to fields selected by a human operator to correspond with particular areas of interest in the test.

When a user selects one of the archived items listed on the electronic index, step 1006, the server subsystem electronically creates at display time to the user an integrated electronic representation of the completed test, step 1006, prepared by the test taker by recompiling electronically the electronic representation of the test form template with the electronic images of the segregated test question responses. The user

is then able to observe an electronic representation of the test taken by the test taker, typically on a computer monitor.

In an alternate embodiment, the recompiling electronically of the electronic representation of the test form template with the electronic images of the segregated test question responses occurs prior to storage on the archival medium.

5 Next, the recompilation step is often accomplished in the same server in which the electronic images of the segregated test question responses and test form template are stored. Optionally, recompilation occurs in a different server computer or even in a client subsystem. An image compression algorithm, such as JPEG, can be

10 applied to the recompiled integrated electronic representation of the completed test prepared by the test taker. The integrated electronic representation of the completed test prepared by the test taker is then stored, in a storage media suitable for archival purposes such as magneto-optical disks, optical disks and other storage media known to those in the art. An electronic index is generated next,

15 where the electronic index can access the integrated electronic representation of the completed test stored on the archival medium. The index is typically a relational database as described above.

When a user selects one of the electronic index selections, step 1020, the integrated electronic representation of the completed test is accessed and may be

20 presented to a human observer using a computer monitor. A browser installed on the computer coupled to the monitor viewed by the human observer will be able to display integrated electronic representation of the completed test and facilitate navigation from one page to the next if the test has multiple pages.

What is Claimed is:

1. In a computerized test processing workflow system configured for analyzing electronic images of test question responses provided by a test taker, the system comprising a server subsystem having at least one server computer and a client subsystem having one or more client computers used by human editors located at the client computers, each human editor having particular skills for examining particular test portion subject matter, a method for enhancing the speed and accuracy of data within the workflow system in order to effect productive use of geographically-distributed human editors having particular appropriate skills, the method comprising:
 - a. selecting from each electronic image of test question responses at least one image clip;
 - b. applying image recognition to generate data electronically from the at least one image clip;
 - c. determining whether verification is needed of the data derived from the at least one image clip;
 - d. matching image clip information to the particular skills of the human editors in order to select human editors having the particular appropriate skills;
 - e. recompiling selected image clips into coherent image portions of the test question responses; and
 - f. selectively presenting the coherent image portions to appropriate human editors located at the remote locations;

whereby, through the transfer of the recompiled coherent image portions to appropriate human editors, widely-distributed human editors of greatly varying skills can be effectively used at the remote locations without the inefficiency of transferring relatively-large, unclipped electronic images of test question responses.
2. The method of claim 1 further comprising the step of analyzing through observation by an appropriate human editor the portion of the response disclosed by the coherent image portions.

3. The method of claim 2 further comprising the step of electronically recording a decision as to the human editors analysis, the decision being generated by the selected human editor for each distributed coherent image portion.
4. The method of claim 2, wherein the step of analyzing the coherent image portions through observation further comprises determining whether the coherent image portion presented matches particular data simultaneously presented to the human editor with the coherent image portion.
5. The method of claim 2, wherein the step of analyzing the coherent image portions through observation further comprises reaching a decision as to what the portion of the response disclosed by the coherent image portion represents.
6. The method of claim 1, wherein the step of applying image recognition to the at least one image clip comprises applying image character recognition or optical character recognition.
7. The method of claim 1, wherein the step of matching image clip information to the particular skills of the human editors further comprises matching the particular test subject matter of the test question with the particular subject matter skills of the human editor.
8. The method of claim 1, wherein the step of matching image clip information to the particular skills of the human editors further comprises using a bit mask of up to 32 bits in which to store information related to each image clip.
9. The method of claim 1 wherein, following the step of selecting at least one image clip from the electronic images of test question responses, the method further comprises:

a. electronically assigning an individual digital identifier to each of the image clips, wherein each individual digital identifier represents at least a portion of information related to the test taker or the test;

b. the step of matching image clip information to the particular skills of the human editors comprises processing the individual digital identifiers related to each coherent image portion against the particular skills of the human editors;

whereby, through the transfer of the recompiled image portions tagged with the digital codes to appropriate human editors, widely-distributed human editors of greatly varying skills can be effectively used at the remote locations without the inefficiency of transferring relatively-large, unclipped electronic images of test question responses.

10. The method of claim 1, further comprising the following steps:

a. recompiling again selected image clips already viewed by appropriate human editors into coherent image portions of the test question response;

b. selectively presenting the coherent image portions to appropriate human editors located at the remote locations but not to the same human editors who previously viewed the coherent image portions; and

c. electronically recording at least one entry generated by the appropriate human editor for each presented coherent image portion.

11. In a computerized test processing workflow system configured for analyzing electronic images of test question responses provided by a test taker, the system comprising a server subsystem having at least one server computer and a client subsystem having one or more client computers used by human editors located at the client computers, each human editor having particular skills for examining particular test portion subject matter, a method for enhancing the speed and accuracy of data within the workflow system in order to effect productive use of geographically-distributed human editors having particular appropriate skills, the method comprising:

a. selecting from each electronic image of test question responses at least one image clip;

- b. applying image recognition to generate data electronically from the at least one image clip;
- c. determining whether verification is needed of the data derived from the at least one image clip;
- d. matching image clip information to the particular skills of the human editors in order to select human editors having the particular appropriate skills;
- e. recompiling selected image clips into coherent image portions of the test question responses;
- f. selectively presenting the coherent image portions to appropriate human editors located at the remote locations;
- g. analyzing through observation by an appropriate human editor the portion of the response disclosed by the coherent image portions;
- h. electronically recording a decision as to the human editors analysis, the decision being generated by the selected human editor for each distributed coherent image portion;

whereby, through the transfer of the recompiled coherent image portions to appropriate human editors, widely-distributed human editors of greatly varying skills can be effectively used at the remote locations without the inefficiency of transferring relatively-large, unclipped electronic images of test question responses.

12. In a computerized test processing workflow system configured for analyzing electronic images of test question responses provided by a test taker, the system comprising a server subsystem having at least one server computer and a client subsystem having one or more client computers used by human editors located at the client computers, each human editor having particular skills for examining particular test portion subject matter, a method for enhancing the speed and accuracy of data within the workflow system in order to effect productive use of geographically-distributed human editors having particular appropriate skills, the method comprising:

- a. selecting from each electronic image of test question responses at least one image clip;

- b. electronically assigning an individual digital identifier to each of the image clips, wherein each individual digital identifier represents at least a portion of information related to the test taker or the test;
- c. applying image recognition to generate data electronically from the at least one image clip;
- d. determining whether verification is needed of the data derived from the at least one image clip;
- e. processing the individual digital identifier related to each image clip against the particular skills of the human editors in order to select human editors having the particular appropriate skills;
- f. recompiling selected image clips into coherent image portions of the test question response;
- g. selectively presenting the coherent image portions to appropriate human editors located at the remote locations;
- h. analyzing through observation by an appropriate human editor the portion of the response disclosed by the coherent image portions;
- i. electronically recording a decision verifying the particular data, the decision being generated by the selected human editor for each distributed coherent image portion;

whereby, through the transfer of the recompiled coherent image portions tagged with the digital codes to appropriate human editors, widely-distributed human editors of greatly varying skills can be effectively used at the remote locations without the inefficiency of transferring for viewing relatively-large, unclipped electronic images of test question responses.

13. The method of claim 12, wherein the step of analyzing through observation includes determining whether the coherent image portion presented matches the particular data, also presented, electronically created from the image clips that form the coherent image portion.

14. The method of claim 12, wherein the step of analyzing through observation includes reaching a decision as to what the portion of the response disclosed by the coherent image portion represents.

15. A system for improving the productive use of geographically-distributed human editors having particular appropriate skills by enhancing the speed and accuracy of data within a computerized test processing workflow system configured for analyzing electronic images of test question responses provided by a test taker, the system comprising

- a. a server subsystem having at least one server computer;
- b. a client subsystem having at least one client computer used by human editors located at the client computers and coupled to the server subsystem;
- c. the at least one client computer configured for selecting from each electronic image of test question responses at least one image clip;
- d. the at least one client computer configured to apply image recognition to generate data electronically from the at least one image clip;
- e. the at least one server computer configured to match image clip information to the particular skills of the human editors in order to select human editors having the particular appropriate skills;
- f. the at least one server computer configured to recompile selected image clips into coherent image portions of the test question responses; and
- g. the at least one client computer configured to selectively present the coherent image portions to appropriate human editors located at the remote locations.

16. The system of claim 15, further comprising the client subsystem configured to facilitate analyzing through observation by an appropriate human editor the portion of the response disclosed by the coherent image portions.

17. The system of claim 15, further comprising the client subsystem configured to facilitate electronically recording a decision verifying the particular data, the decision

being generated by the selected human editor for each distributed coherent image portion.

18. The system of claim 15, further comprising the server subsystem configured to electronically assign an individual digital identifier to each of the image clips, wherein each individual digital identifier represents at least a portion of information related to the test taker or the test.

19. The system of claim 18, further comprising the server subsystem configured to process the individual digital identifier related to each image clip against the particular skills of the human editors in order to select human editors having the particular appropriate skills.

20. The system of claim 15, further comprising the server subsystem configured to:

- a. recompile again selected image clips already viewed by appropriate human editors into coherent image portions of the test question response;
- b. selectively present the coherent image portions to appropriate human editors located at the remote locations but not to the same human editors who previously viewed the coherent image portions; and
- c. electronically record at least one entry generated by the appropriate human editor for each presented coherent image portion.

21. A computer program product readable by a computing system for a computer process for a computerized test processing workflow system configured for analyzing electronic images of test question responses provided by a test taker, the system comprising a server subsystem having at least one server computer and a client subsystem having one or more client computers used by human editors located at the client computers, each human editor having particular skills for examining particular test portion subject matter, a method for enhancing the speed and accuracy of data within the workflow system in order to effect productive use of geographically-distributed human editors having particular appropriate skills, the computer program

product comprising encoded instructions either embedded on a computer-readable media or transmitted in a propagating wave, the encoded instructions being arranged and configured to be read by the computing system for causing the computing system to perform the steps of:

- a. selecting from each electronic image of test question responses at least one image clip;
- b. applying image recognition to generate data electronically from the at least one image clip;
- c. determining whether verification is needed of the data derived from the at least one image clip;
- d. matching image clip information to the particular skills of the human editors in order to select human editors having the particular appropriate skills;
- e. recompiling selected image clips into coherent image portions of the test question responses; and
- f. selectively presenting the coherent image portions to appropriate human editors located at the remote locations;

whereby, through the transfer of the recompiled coherent image portions to appropriate human editors, widely-distributed human editors of greatly varying skills can be effectively used at the remote locations without the inefficiency of transferring relatively-large, unclipped electronic images of test question responses.

22. A computer program product readable by a computing system for a computer process for a computerized test processing workflow system configured for analyzing electronic images of test question responses provided by a test taker, the system comprising a server subsystem having at least one server computer and a client subsystem having one or more client computers used by human editors located at the client computers, each human editor having particular skills for examining particular test portion subject matter, a method for enhancing the speed and accuracy of data within the workflow system in order to effect productive use of geographically-distributed human editors having particular appropriate skills, the computer program product comprising encoded instructions either embedded on a computer-readable

media or transmitted in a propagating wave, the encoded instructions being arranged and configured to be read by the computing system for causing the computing system to perform the steps of:

- a. selecting from each electronic image of test question responses at least one image clip;
- b. applying image recognition to generate data electronically from the at least one image clip;
- c. determining whether verification is needed of the data derived from the at least one image clip;
- d. matching image clip information to the particular skills of the human editors in order to select human editors having the particular appropriate skills;
- e. recompiling selected image clips into coherent image portions of the test question responses;
- f. selectively presenting the coherent image portions to appropriate human editors located at the remote locations;
- g. analyzing through observation by an appropriate human editor the portion of the response disclosed by the coherent image portions;
- h. electronically recording a decision as to the human editors analysis, the decision being generated by the selected human editor for each distributed coherent image portion;

whereby, through the transfer of the recompiled coherent image portions to appropriate human editors, widely-distributed human editors of greatly varying skills can be effectively used at the remote locations without the inefficiency of transferring relatively-large, unclipped electronic images of test question responses.

23. A system for improving the productive use of geographically-distributed human editors having particular appropriate skills by enhancing the speed and accuracy of data within a computerized test processing workflow system configured for analyzing electronic images of test question responses provided by a test taker, the system comprising:

- a. means for selecting from each electronic image of test question responses at least one image clip;
- b. means for applying image recognition to generate data electronically from the at least one image clip;
- c. means for matching image clip information to the particular skills of the human editors in order to select human editors having the particular appropriate skills;
- d. means for recompiling selected image clips into coherent image portions of the test question responses; and
- e. means for present the coherent image portions to appropriate human editors located at the remote locations.

FIG. 1

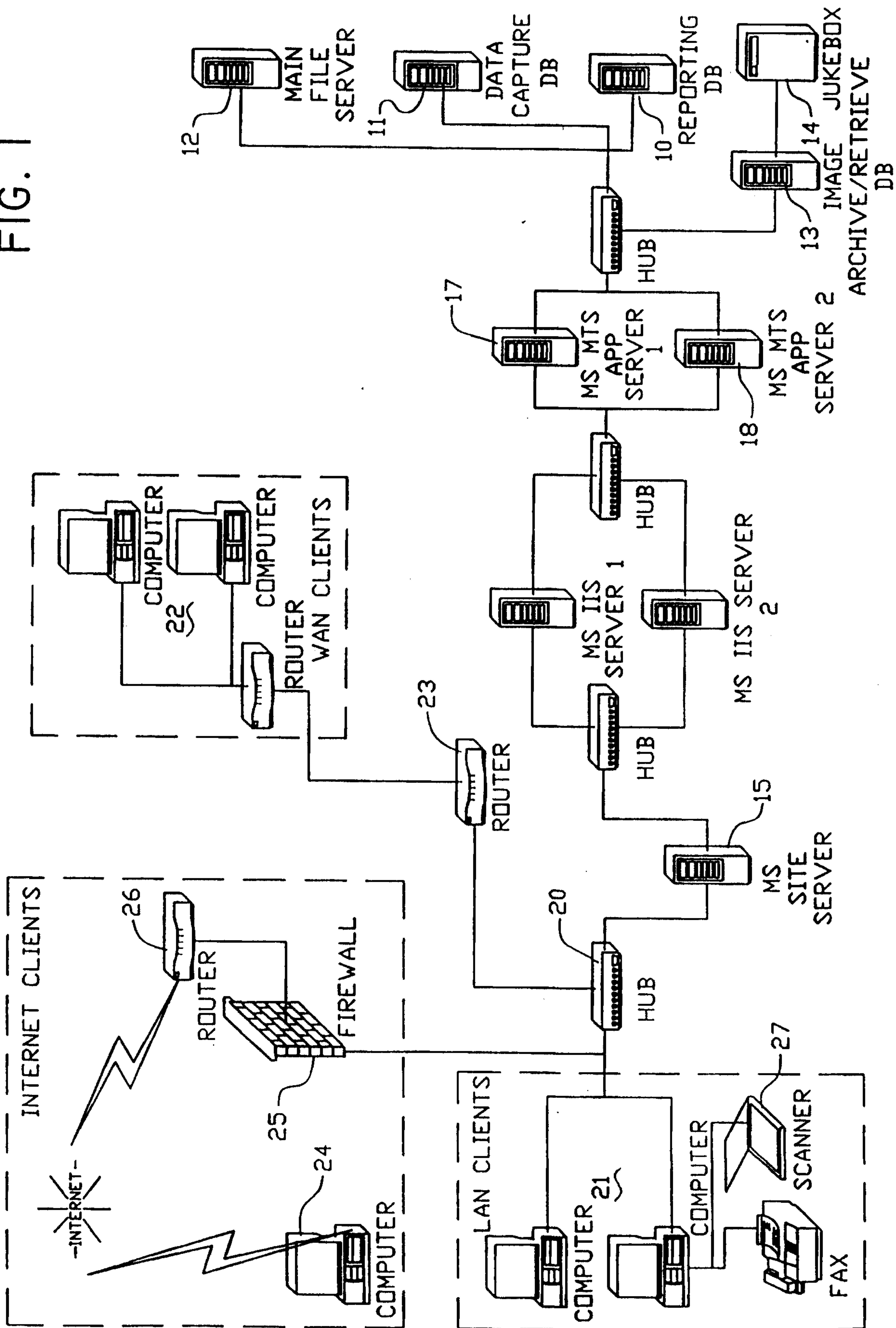
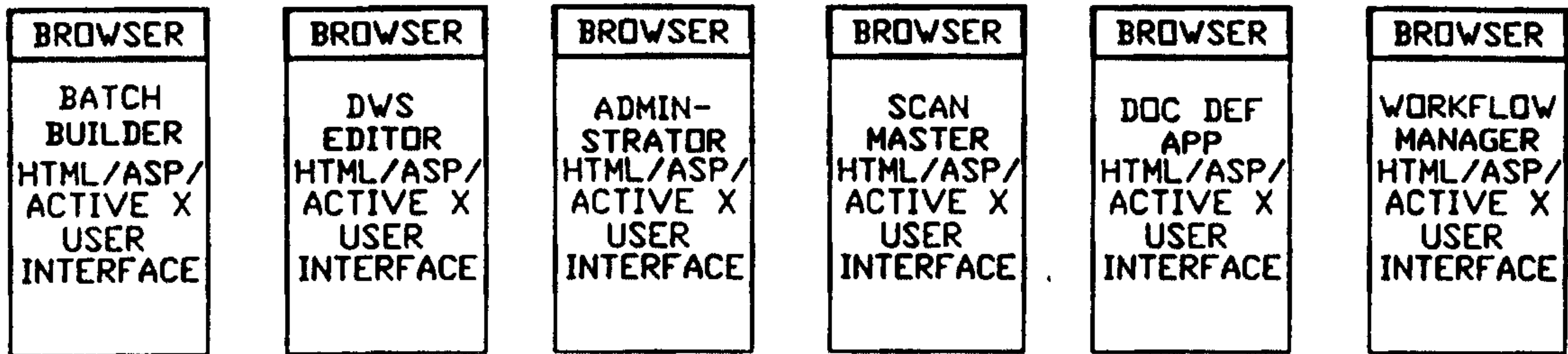
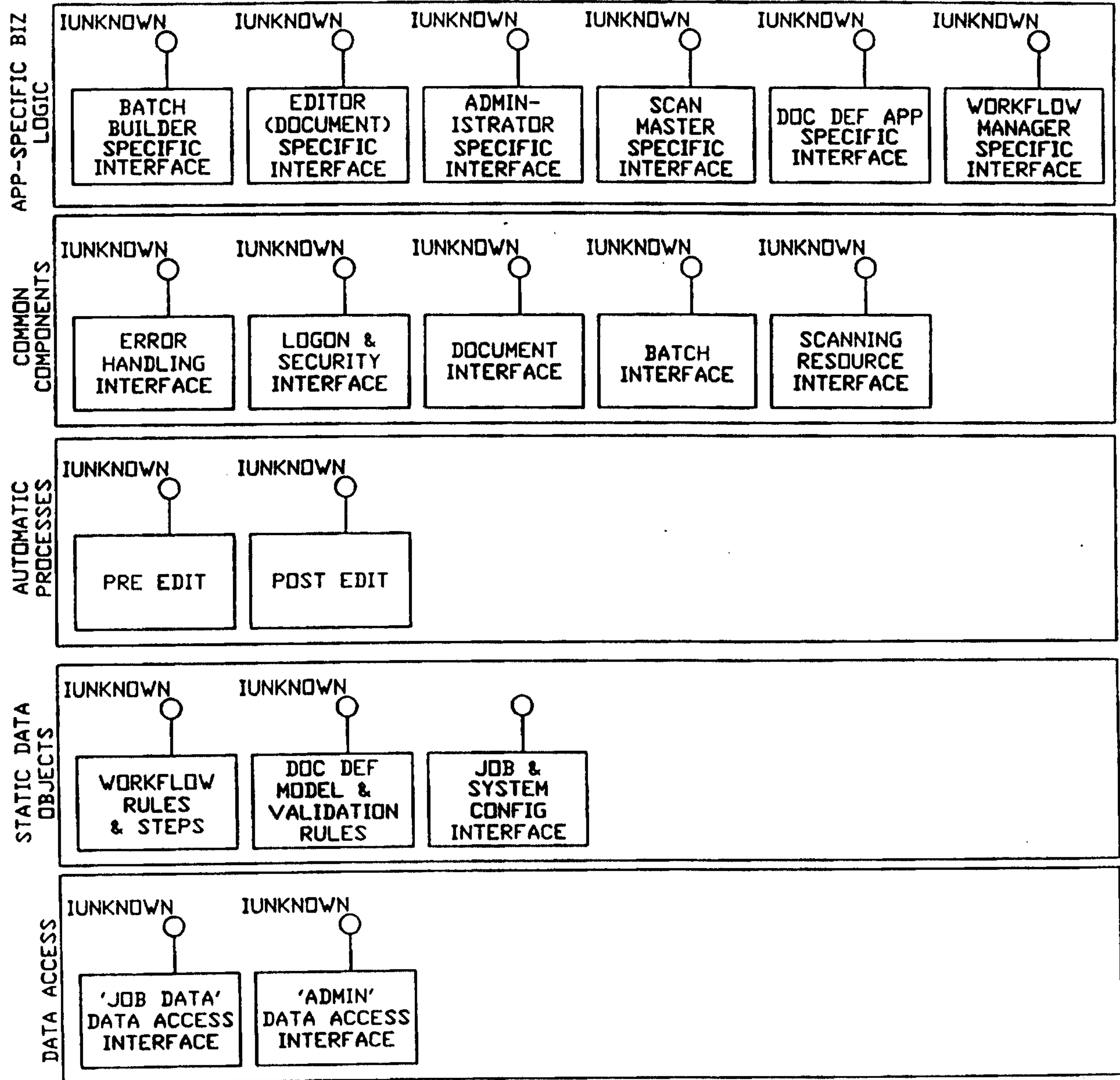


FIG. 1A

Presentation
Layer

BUSINESS OBJECT LAYER



DATA LAYER

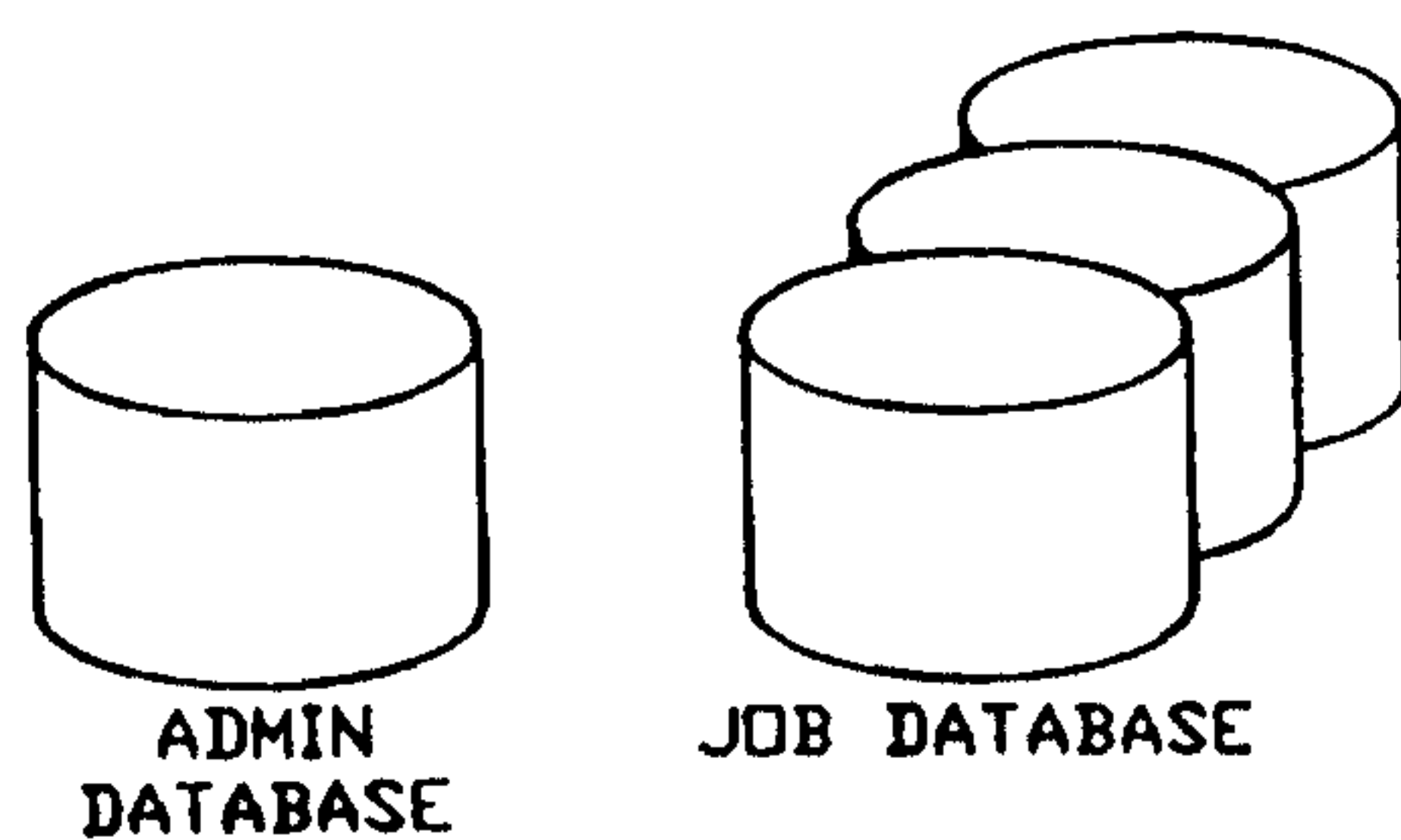


FIG. 2

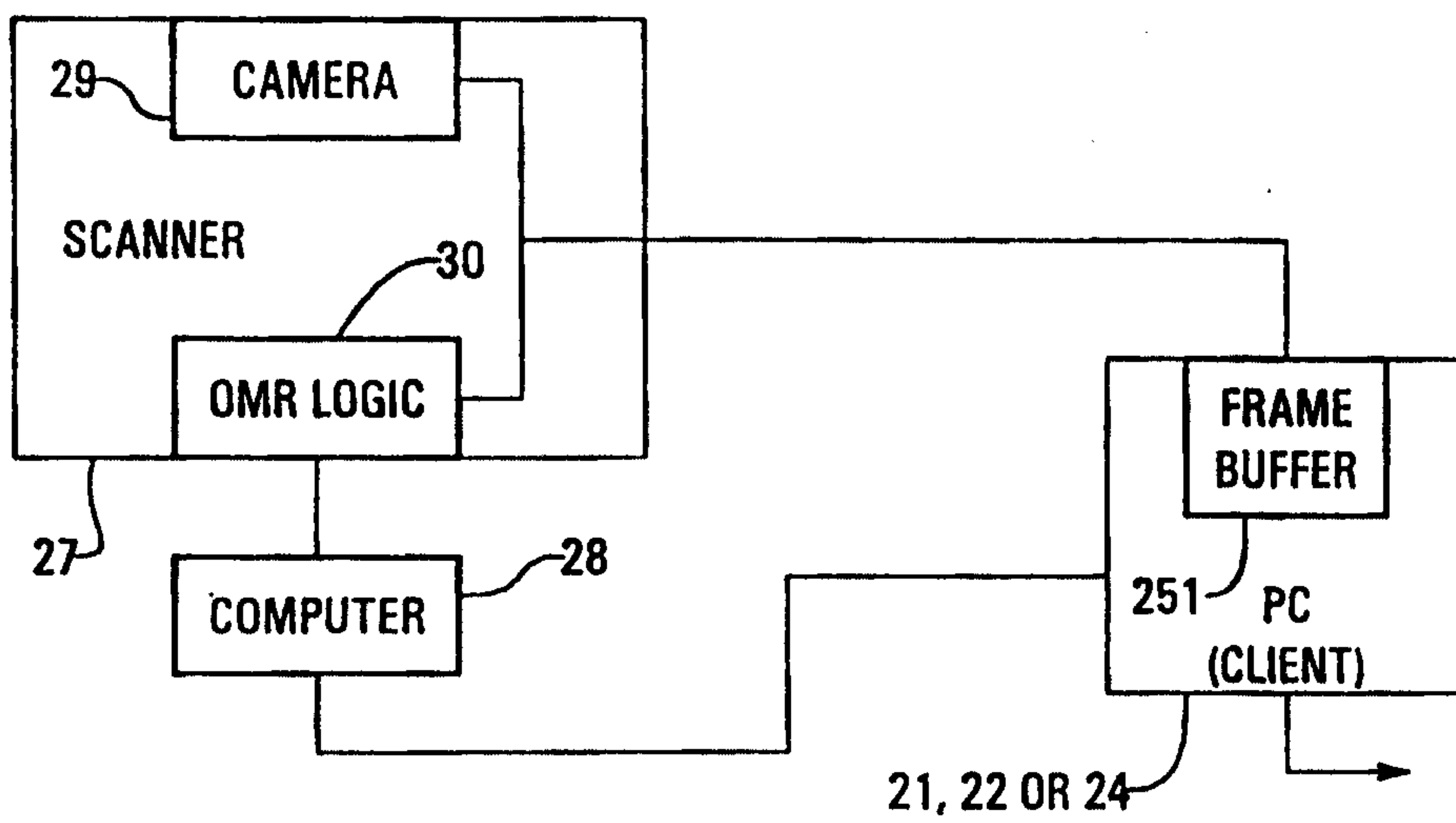


FIG. 3

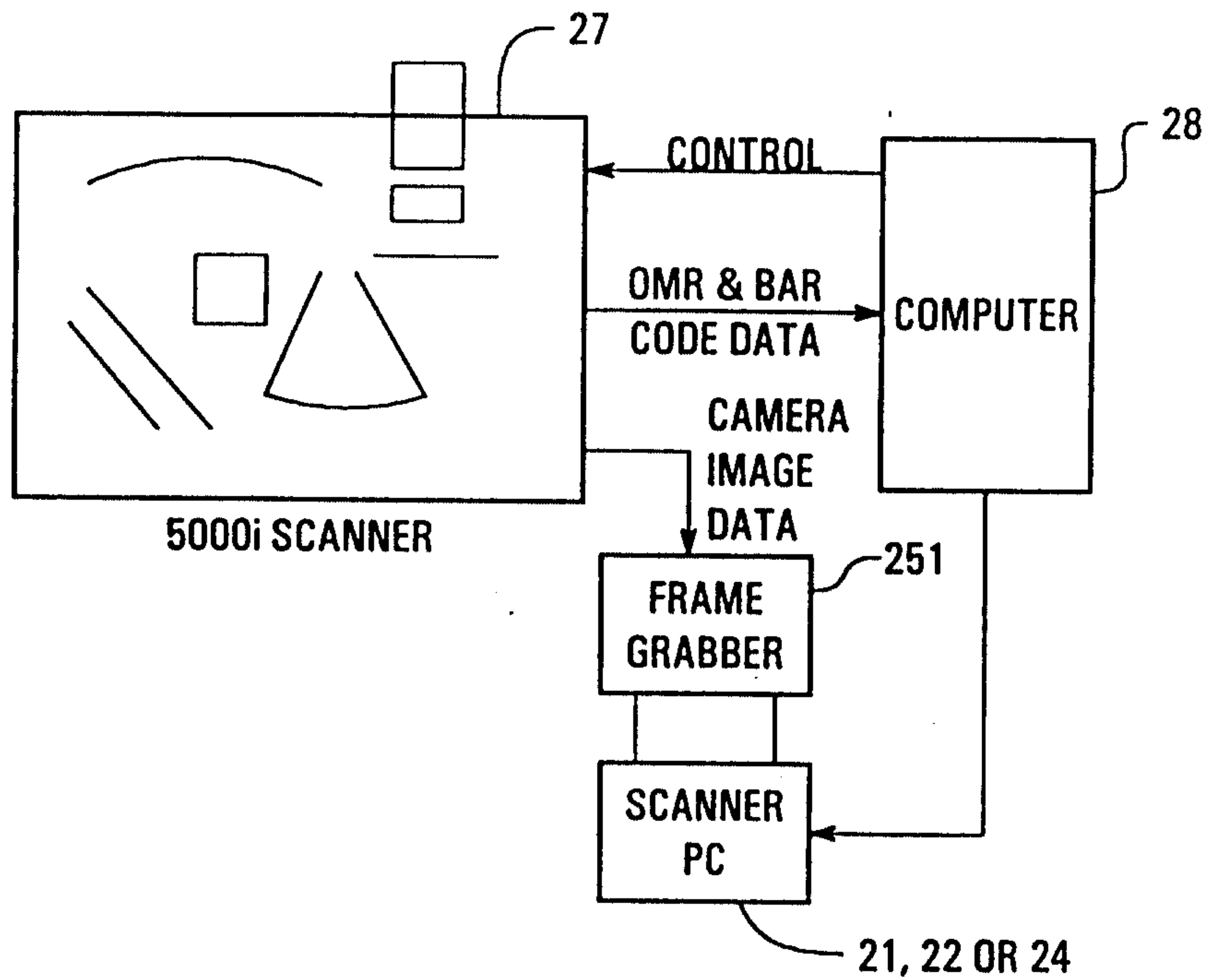


FIG. 4

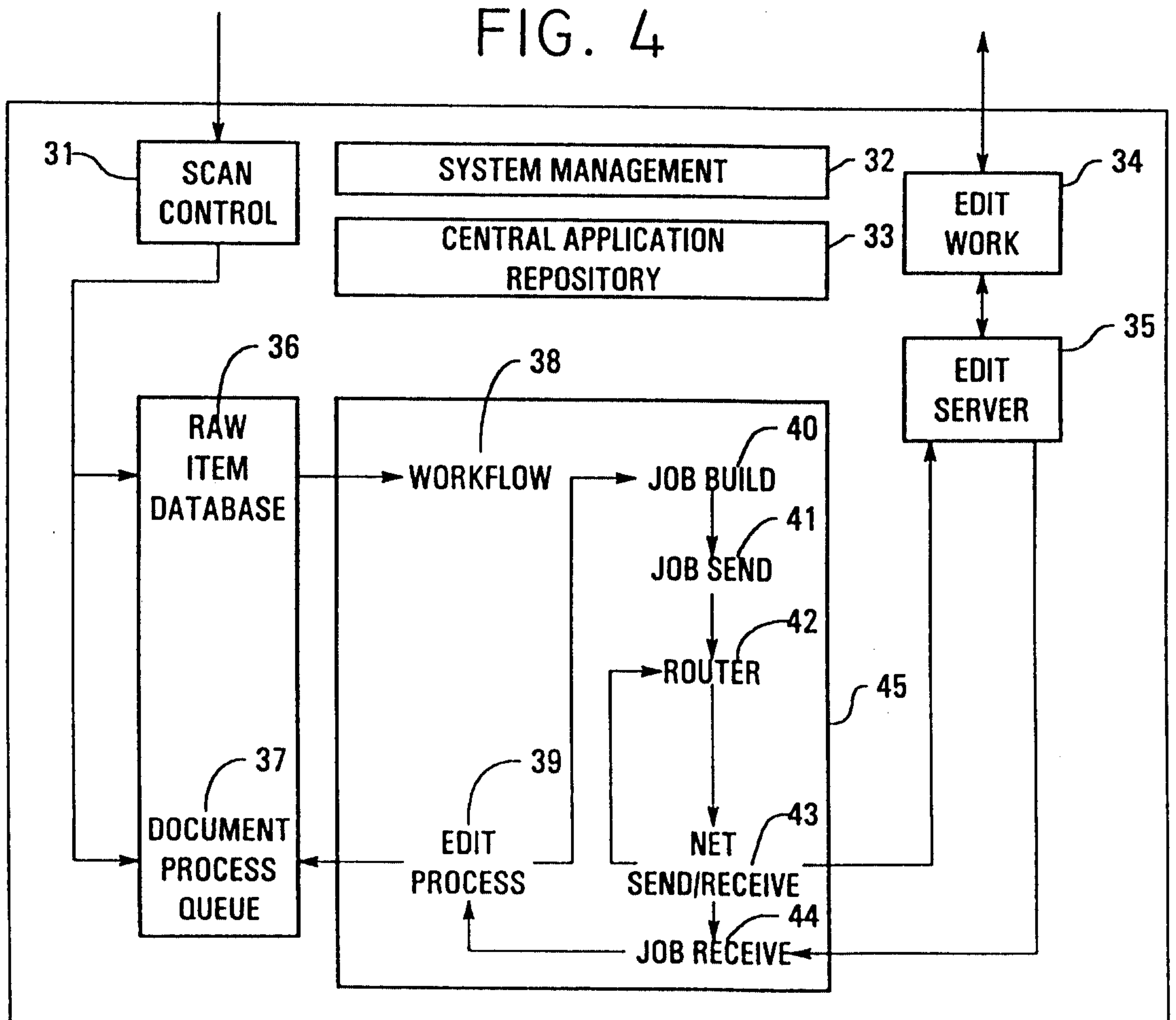


FIG. 5

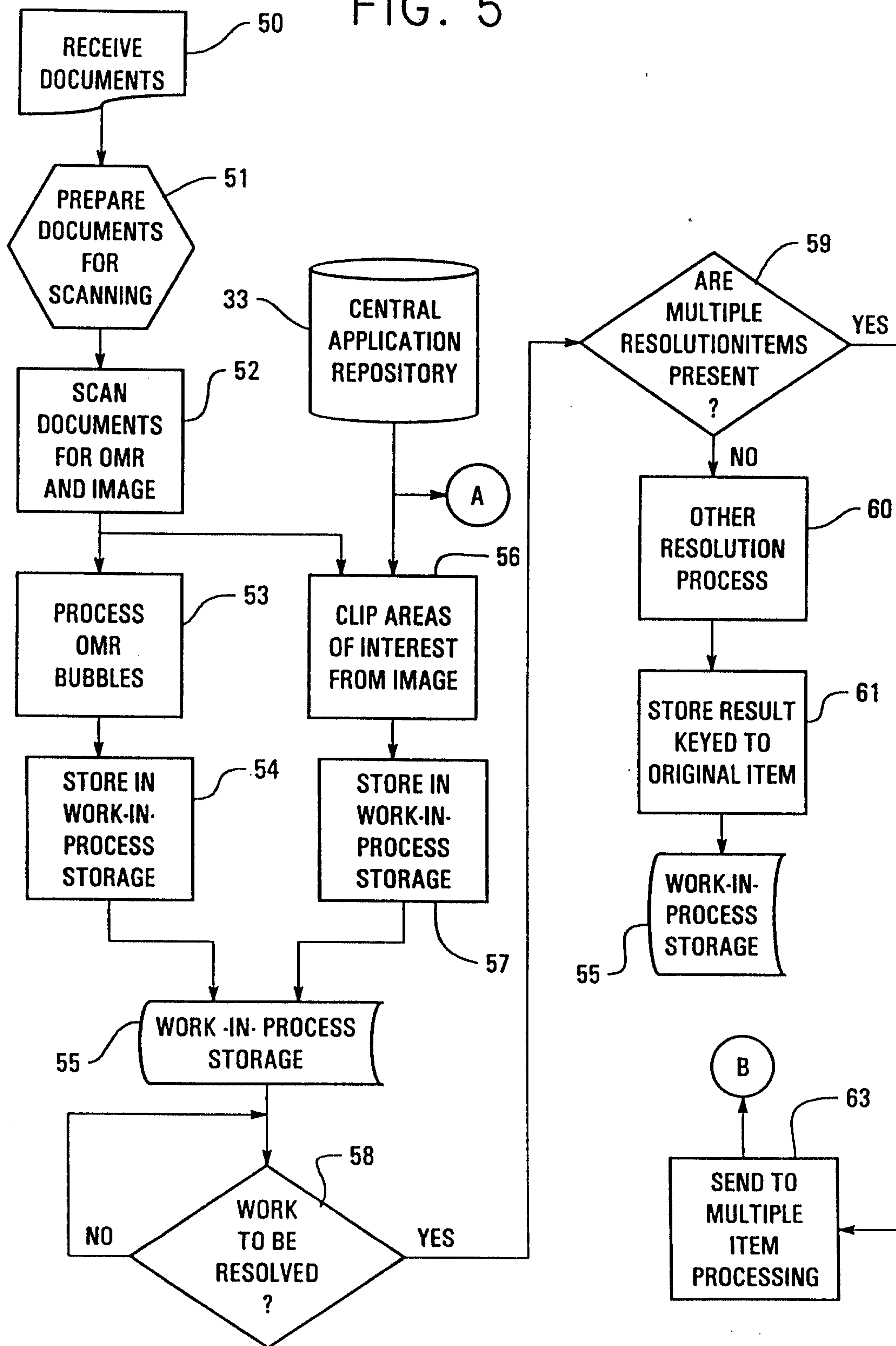


FIG. 6-1

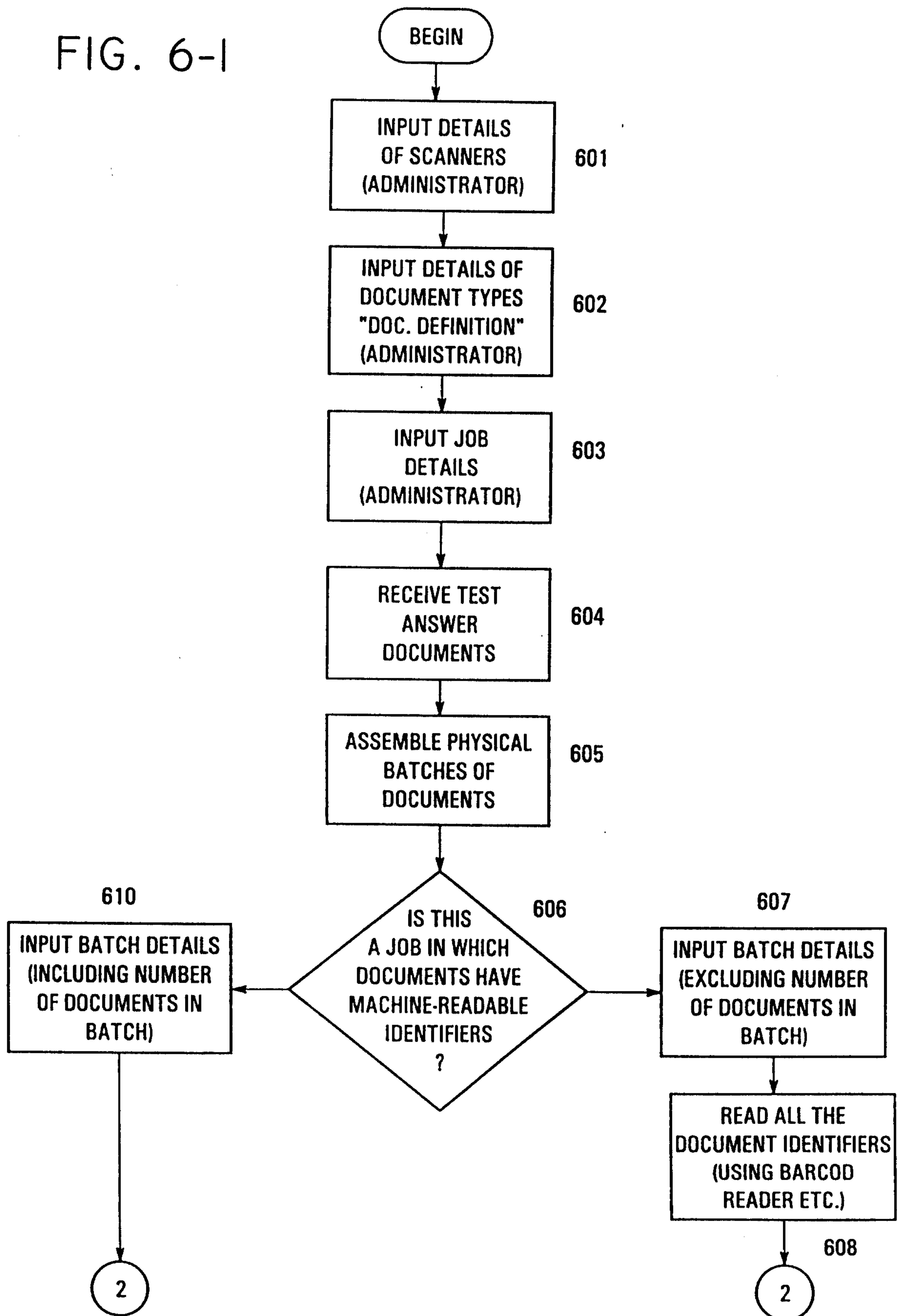


FIG. 6-2

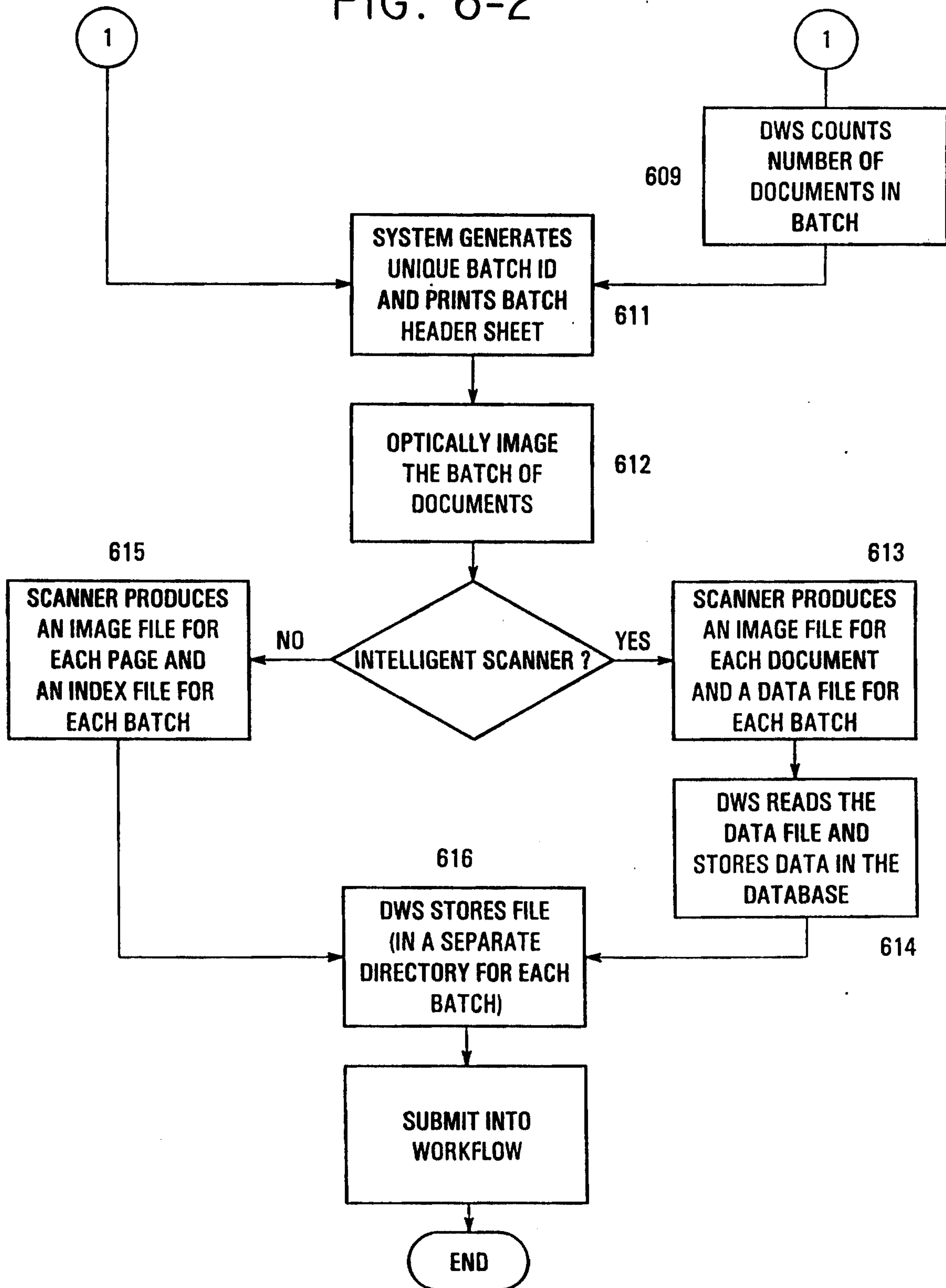


FIG. 7

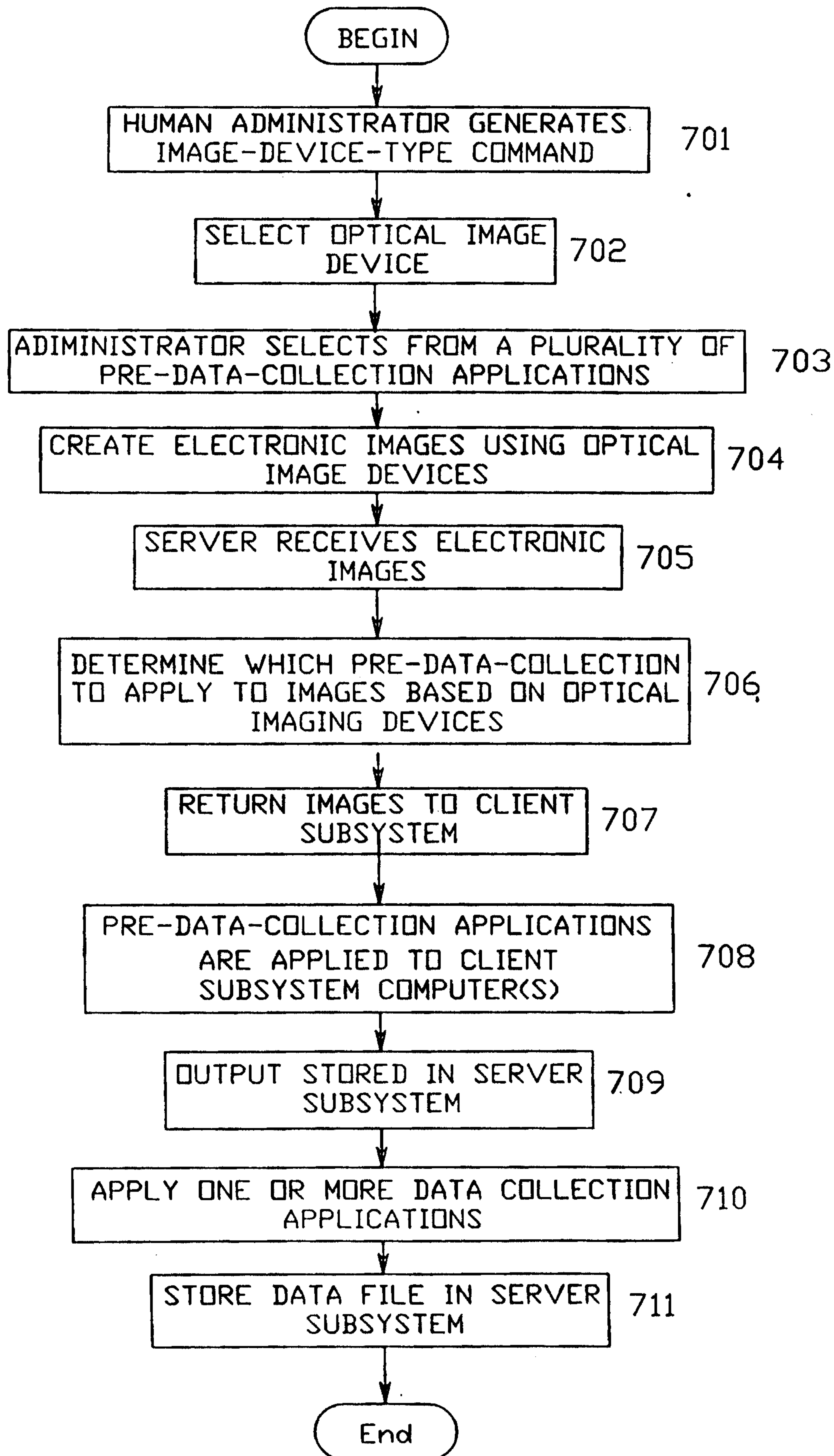


FIG. 8

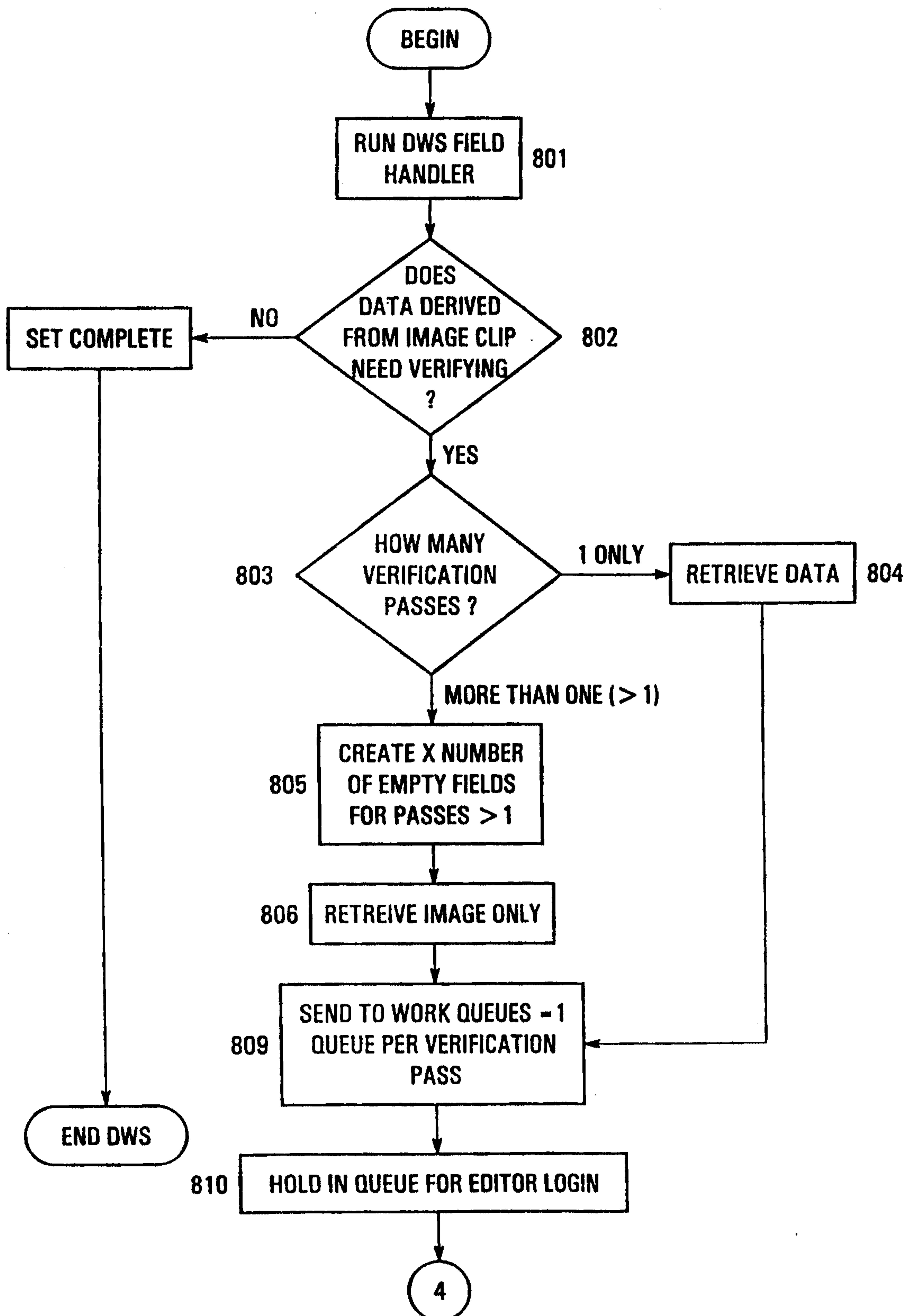


FIG. 8A

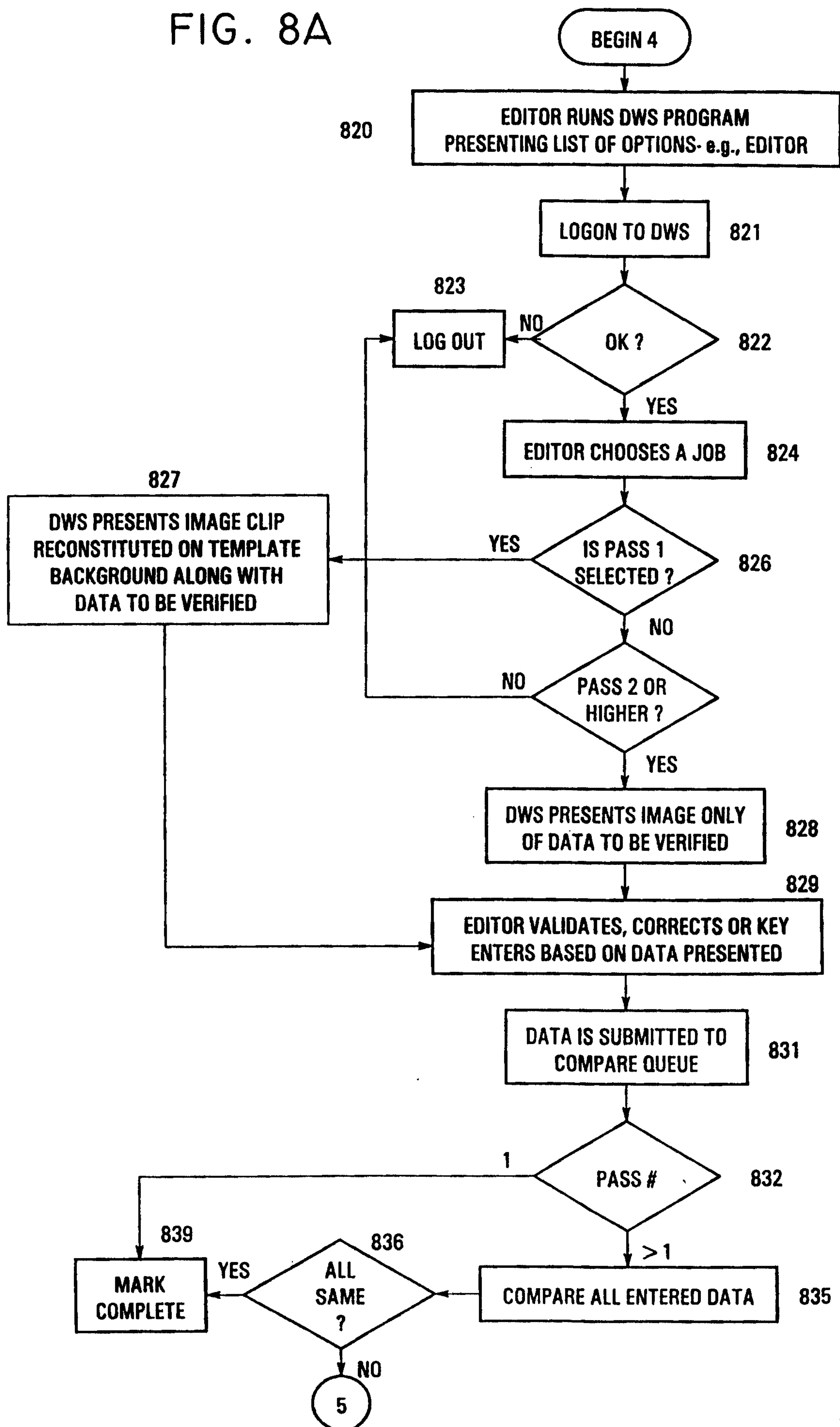


FIG. 8B

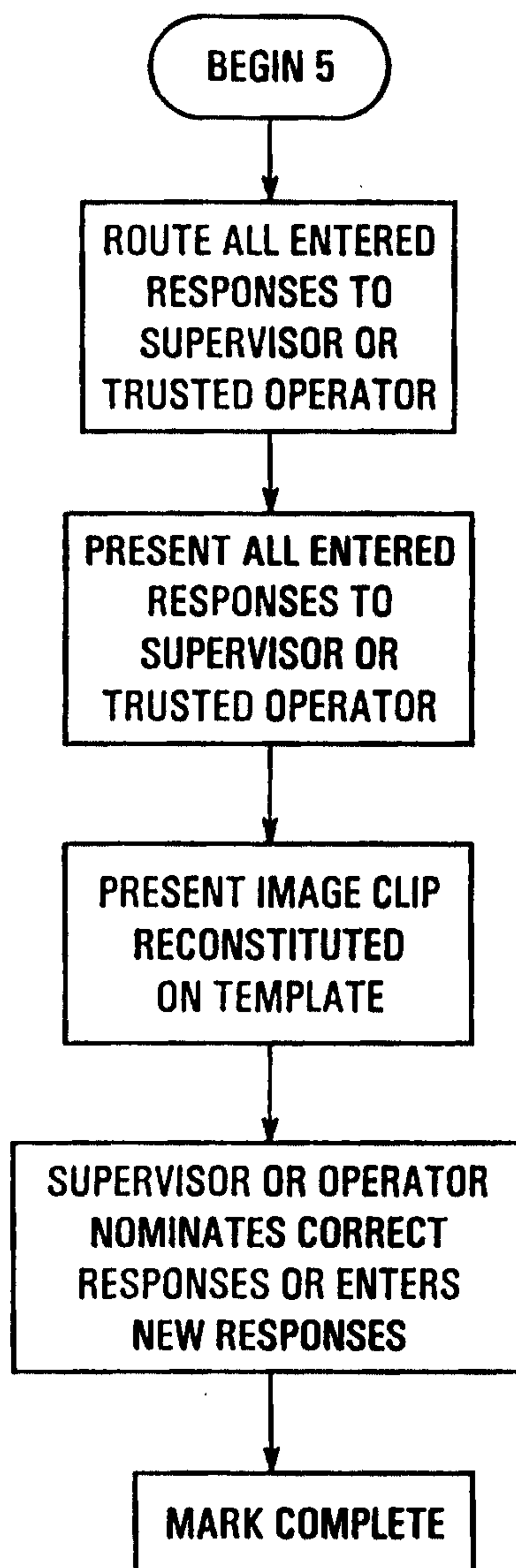


FIG. 9

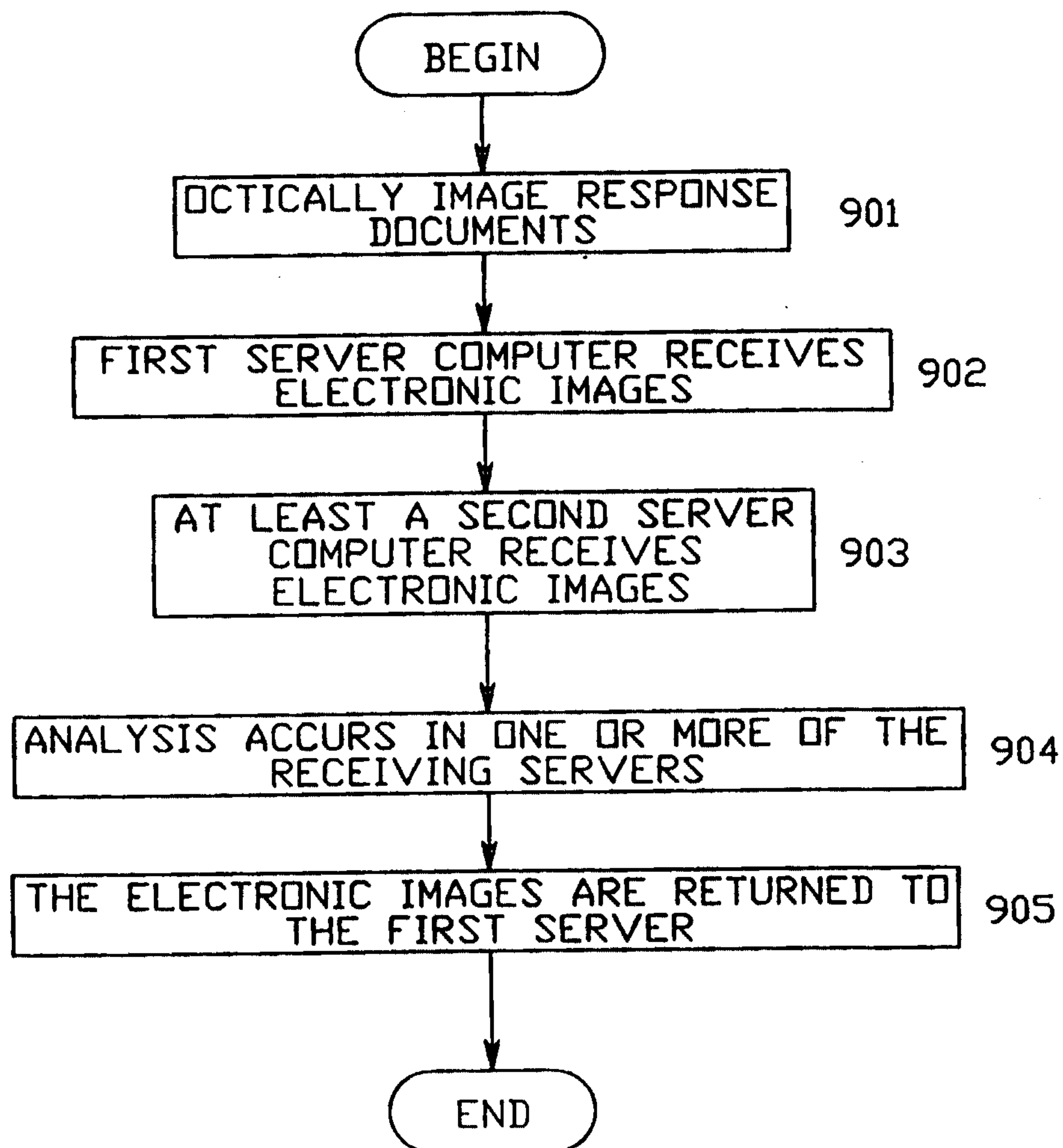


FIG. 9A

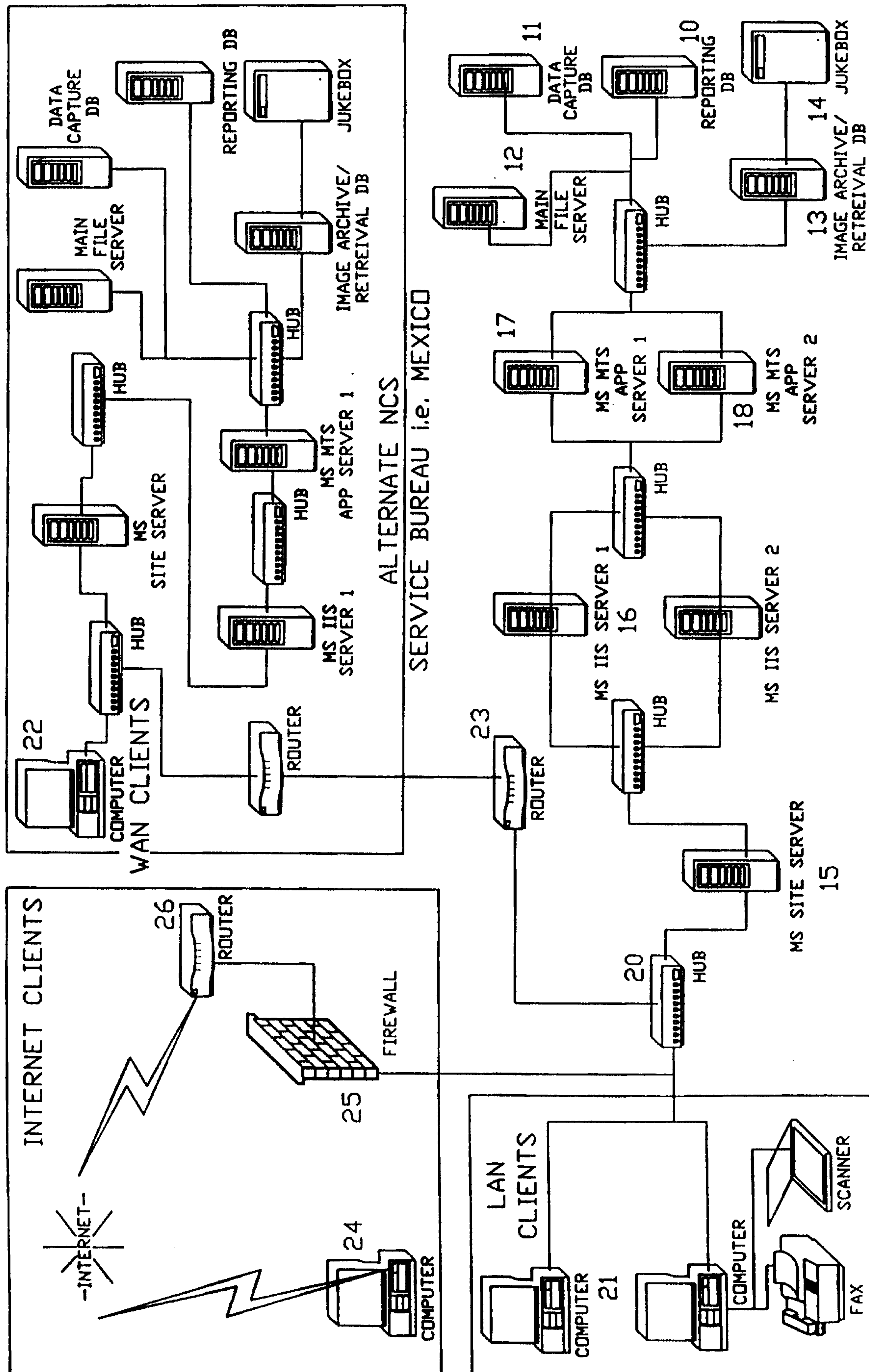


FIG. 10

