(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0134251
(43) 공개일자 2017년12월06일

(51) 국제특허분류(Int. Cl.)
H04N 1/32 (2006.01)
(52) CPC특허분류
H04N 1/32277 (2013.01)
H04N 1/32288 (2013.01)
(21) 출원번호 10-2017-0065128
(22) 출원일자 2017년05월26일
심사청구일자 없음
(30) 우선권주장
JP-P-2016-106494 2016년05월27일 일본(JP)

(71) 출원인
캐논 가부시끼가이샤
일본 도쿄도 오오따꾸 시모마루꼬 3쪼메 30방 2고
(72) 발명자
나카츠카 다다노리
일본 도쿄도 오오따꾸 시모마루꼬 3쪼메 30방 2고 캐논 가부시끼가이샤 내
(74) 대리인
장수길, 이중희

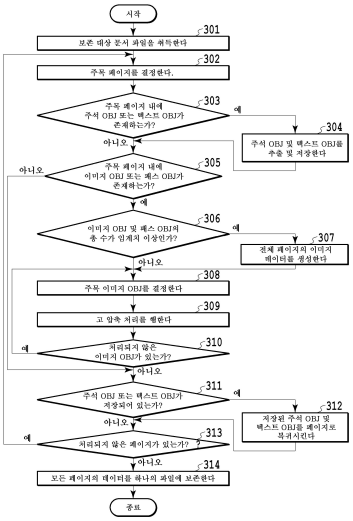
전체 청구항 수 : 총 19 항

(54) 발명의 명칭 **문서 파일의 압축 처리를 행하는 화상 처리 장치, 문서 파일의 압축 방법 및 저장 매체**

(57) 요 약

페이지 내에 주석 및 텍스트를 포함하는 문서 파일의 경우에도, 주석 및 텍스트 정보를 유지하면서 파일 사이즈가 감소된다. 화상 처리 장치는 입력된 문서 파일의 페이지로부터 주석 오브젝트 및 텍스트 오브젝트를 추출하고, 상기 추출된 주석 오브젝트 및 텍스트 오브젝트를 제외한 상기 페이지 내에 포함되는 다른 오브젝트를 1개의 이미지 오브젝트로 변환하고, 상기 변환된 1개의 이미지 오브젝트를 압축하며, 상기 압축된 이미지 오브젝트와 상기 추출된 상기 주석 오브젝트 및 상기 텍스트 오브젝트에 기초하여 페이지 데이터를 생성한다.

대 표 도 - 도3



명 세 서

청구범위

청구항 1

화상 처리 장치이며,

입력된 문서 파일의 페이지로부터 주석 오브젝트 및 텍스트 오브젝트를 추출하도록 구성되는 추출 유닛과,

상기 추출된 주석 오브젝트 및 텍스트 오브젝트를 제외한 상기 페이지 내에 포함되는 다른 오브젝트를 1개의 이미지 오브젝트로 변환하고, 상기 변환된 1개의 이미지 오브젝트를 압축하도록 구성되는 압축 유닛과,

상기 압축 유닛에 의해 압축된 이미지 오브젝트와, 상기 추출 유닛에 의해 추출된 상기 주석 오브젝트 및 상기 텍스트 오브젝트에 기초하여, 페이지 데이터를 생성하도록 구성되는 생성 유닛을 포함하는, 화상 처리 장치.

청구항 2

제1항에 있어서,

상기 압축 유닛은, 상기 추출된 주석 오브젝트 및 텍스트 오브젝트를 제외한 상기 페이지 내에 포함되는 다른 오브젝트의 수가 미리결정된 임계치 이상인 경우, 상기 다른 오브젝트를 1개의 이미지 오브젝트로 변환하고, 상기 변환된 1개의 이미지 오브젝트를 압축하는, 화상 처리 장치.

청구항 3

제1항에 있어서,

상기 추출된 주석 오브젝트 및 텍스트 오브젝트를 제외한 상기 페이지 내에 포함된 상기 다른 오브젝트는 이미지 오브젝트 또는 패스(path) 오브젝트인, 화상 처리 장치.

청구항 4

제3항에 있어서,

상기 압축 유닛은, 상기 페이지 내에 포함되는 이미지 오브젝트와 패스 오브젝트의 총 수가 상기 미리결정된 임계치 이상인 경우, 상기 페이지 내에 포함되는 상기 이미지 오브젝트와 상기 패스 오브젝트를 1개의 이미지 오브젝트로 변환하고, 상기 변환된 1개의 이미지 오브젝트를 압축하는, 화상 처리 장치.

청구항 5

제3항에 있어서,

상기 미리결정된 임계치로서, 상기 이미지 오브젝트의 수와 비교되는 제1 임계치와, 상기 패스 오브젝트의 수와 비교되는 제2 임계치가 별개로 설정되고,

상기 압축 유닛은, 상기 이미지 오브젝트의 수가 상기 제1 임계치 이상인 경우 및 상기 패스 오브젝트의 수가 상기 제2 임계치 이상인 경우 중 적어도 하나의 경우에, 상기 페이지 내에 포함되는 상기 이미지 오브젝트와 상기 패스 오브젝트를 1개의 이미지의 오브젝트로 변환하고, 상기 변환된 1개의 이미지 오브젝트를 압축하는, 화상 처리 장치.

청구항 6

제2항에 있어서,

상기 압축 유닛은, 상기 추출된 주석 오브젝트 및 텍스트 오브젝트를 제외한 상기 페이지 내에 포함되는 다른 오브젝트의 수가 상기 미리결정된 임계치보다 작은 경우, 상기 다른 오브젝트를 1개의 이미지 오브젝트로 변환하지 않는, 화상 처리 장치.

청구항 7

제6항에 있어서,

상기 압축 유닛은, 상기 추출된 주석 오브젝트 및 텍스트 오브젝트를 제외한 상기 페이지에 포함되는 다른 오브젝트의 수가 상기 미리결정된 임계치보다 작은 경우, 상기 페이지 내에 포함되는 각각의 개별 이미지 오브젝트를 압축하며,

상기 생성 유닛은, 상기 페이지 내에 포함되는 상기 압축된 각각의 이미지 오브젝트, 상기 페이지 내에 포함되는 패스 오브젝트, 및 상기 추출된 주석 오브젝트 및 텍스트 오브젝트에 기초하여 페이지 데이터를 생성하는, 화상 처리 장치.

청구항 8

제1항에 있어서,

상기 압축 유닛은, 상기 변환된 1개의 이미지 오브젝트를 텍스트 영역 및 배경 영역으로 변환하며 상기 텍스트 영역 및 배경 영역에 각각 상이한 방식의 압축 처리를 적용하는, 화상 처리 장치.

청구항 9

제8항에 있어서,

상기 압축 유닛은, 상기 텍스트 영역에 포함되는 문자 화상에는 가역 방식의 압축 처리를 적용하고 상기 배경 영역에 포함되는 배경 화상에는 비가역 방식의 압축 처리를 적용하는, 화상 처리 장치.

청구항 10

제1항에 있어서,

상기 압축 유닛은, 이미지 오브젝트의 리소스가 상기 입력된 문서 파일에 포함되는 복수의 페이지에 의해 공유되는 경우, 상기 추출된 주석 오브젝트 및 텍스트 오브젝트를 제외한 상기 페이지 내에 포함되는 다른 오브젝트 및 상기 복수의 페이지에 의해 공유되는 이미지 오브젝트의 리소스를 1개의 이미지 오브젝트로 변환하고, 상기 변환된 1개의 이미지 오브젝트를 압축하는, 화상 처리 장치.

청구항 11

제1항에 있어서,

상기 추출 유닛은, 불투명한 색으로 채워지도록 지정된 오브젝트가 상기 입력된 문서 파일의 페이지의 주석 오브젝트 또는 텍스트 오브젝트와 겹치는 경우, 상기 주석 오브젝트 및 겹쳐 있는 오브젝트 또는 상기 텍스트 오브젝트 및 겹쳐 있는 오브젝트를 1개의 이미지 오브젝트로 변환하고, 상기 불투명한 색으로 채워지도록 지정된 오브젝트에 의해 겹치지 않는 주석 오브젝트 또는 텍스트 오브젝트를 추출하는, 화상 처리 장치.

청구항 12

화상 처리 장치에서의 문서 파일의 압축 방법이며,

입력된 문서 파일의 페이지로부터 주석 오브젝트 및 텍스트 오브젝트를 추출하는 단계와,

상기 추출된 주석 오브젝트 및 텍스트 오브젝트를 제외한 상기 페이지 내에 포함되는 다른 오브젝트를 1개의 이미지 오브젝트로 변환하는 단계와,

상기 변환된 1개의 이미지 오브젝트를 압축하는 단계와,

상기 압축된 이미지 오브젝트와 상기 추출된 주석 오브젝트 및 텍스트 오브젝트에 기초하여 페이지 데이터를 생성하는 단계를 포함하는, 문서 파일의 압축 방법.

청구항 13

컴퓨터가 방법을 실행하게 하는 프로그램을 저장하는 컴퓨터 판독가능 저장 매체이며,

상기 방법은,

입력된 문서 파일의 페이지로부터 주석 오브젝트 및 텍스트 오브젝트를 추출하는 추출 단계와,

상기 추출된 주석 오브젝트 및 텍스트 오브젝트를 제외한 상기 페이지 내에 포함되는 다른 오브젝트를 1개의 이미지 오브젝트로 변환하는 변환 단계와,

상기 변환된 1개의 이미지 오브젝트를 압축하는 압축 단계와,

상기 압축된 이미지 오브젝트와 상기 추출된 주석 오브젝트 및 텍스트 오브젝트에 기초하여 페이지 데이터를 생성하는 생성 단계를 포함하는, 컴퓨터 판독가능 저장 매체.

청구항 14

제13항에 있어서,

상기 변환 단계에서는, 상기 추출된 주석 오브젝트 및 텍스트 오브젝트를 제외한 상기 페이지 내에 포함되는 상기 다른 오브젝트의 수가 미리결정된 임계치 이상인 경우, 상기 다른 오브젝트가 1개의 이미지 오브젝트로 변환되는, 컴퓨터 판독가능 저장 매체.

청구항 15

제13항에 있어서,

상기 추출된 주석 오브젝트 및 텍스트 오브젝트를 제외한 상기 페이지 내에 포함된 상기 다른 오브젝트는 이미지 오브젝트 또는 패스 오브젝트인, 컴퓨터 판독가능 저장 매체.

청구항 16

제14항에 있어서,

상기 변환 단계에서는, 상기 추출된 주석 오브젝트와 텍스트 오브젝트를 제외한 상기 페이지 내에 포함되는 상기 다른 오브젝트의 수가 상기 미리결정된 임계치보다 작은 경우, 상기 다른 오브젝트가 1개의 이미지 오브젝트로 변환되지 않는, 컴퓨터 판독가능 저장 매체.

청구항 17

제13항에 있어서,

상기 압축 단계에서, 상기 변환된 1개의 이미지 오브젝트는 텍스트 영역 및 배경 영역으로 분할되고, 상이한 방식의 압축 처리가 상기 텍스트 영역 및 상기 배경 영역에 각각 적용되는, 컴퓨터 판독가능 저장 매체.

청구항 18

제13항에 있어서,

상기 변환 단계에서는, 이미지 오브젝트의 리소스가 상기 입력된 문서 파일에 포함되는 복수의 페이지에 의해 공유되는 경우, 상기 추출된 주석 오브젝트 및 텍스트 오브젝트를 제외한 상기 페이지 내에 포함되는 다른 오브젝트 및 상기 복수의 페이지에 의해 공유되는 상기 이미지 오브젝트의 리소스가 1개의 이미지 오브젝트로 변환되는, 컴퓨터 판독가능 저장 매체.

청구항 19

제13항에 있어서,

상기 추출 단계에서는, 불투명한 색으로 채워지도록 지정된 오브젝트가 상기 입력된 문서 파일의 페이지의 상기 주석 오브젝트 또는 텍스트 오브젝트와 겹치는 경우, 상기 주석 오브젝트 및 겹쳐 있는 오브젝트 또는 상기 텍스트 오브젝트 및 상기 겹쳐 있는 오브젝트가 1개의 이미지 오브젝트로 변환되며, 상기 불투명한 색으로 채워지도록 지정된 오브젝트에 의해 겹치지 않는 주석 오브젝트 또는 텍스트 오브젝트가 추출되는, 컴퓨터 판독가능 저장 매체.

발명의 설명

기 술 분 야

[0001] 본 발명은, 문서 데이터의 압축 기술에 관한 것이다.

배 경 기 술

[0002] 문서를 보존하는 기술로서, 그 목적이나 용도에 따라서 종래부터 여러 가지 제안이 이루어져 있다. 예를 들어, 일본 특허 공개 공보 제2000-332984호는, 문서 화상을 보존할 때의 데이터양 억제와 화상 품질 유지의 목적으로, 문서 내의 텍스트 영역과 이미지 영역 사이의 비율에 기초하여, 문서를 텍스트계 문서와 픽처계 문서로 분류함으로써 보존 방법을 변경하는 기술을 제안하고 있다.

[0003] 여기서, 예를 들어 PDF와 같이, 페이지를 구성하는 콘텐츠로서, 이미지 외에, 텍스트, 패스(path)(도형), 및 주석과 같은 오브젝트를 개별적으로 저장할 수 있는 포맷의 문서 파일을 보존하는 경우를 생각한다. 주석이란, 본문을 보충, 강조 등을 할 목적으로 첨부되는, 본문과 독립적인 오브젝트이다. 이러한 포맷의 문서 파일을, 상기 일본 특허 공개 공보 제2000-332984호의 기술을 적용하여 보존하고자 시도하는 경우, 더 이상 페이지의 이미지 데이터의 생성에 동반하여 주석 정보를 유지할 수 없게 된다. 또한, 더 이상 페이지 내의 텍스트 정보 부분의 편집도 할 수 없게 된다. 또한, 문서의 보존 시에, 일반적으로 그 데이터양을 최소한으로 억제하는 것도 요구되지만, 일본 특허 공개 공보 제2000-332984호는 이러한 점을 특히 고려하고 있지 않다.

발명의 내용

과제의 해결 수단

[0004] 본 발명에 따른 화상 처리 장치는, 입력된 문서 파일의 페이지로부터 주석 오브젝트 및 텍스트 오브젝트를 추출하도록 구성되는 추출 유닛과; 상기 추출된 주석 오브젝트 및 텍스트 오브젝트를 제외한 상기 페이지 내에 포함되는 다른 오브젝트를 1개의 이미지 오브젝트로 변환하고, 상기 변환된 1개의 이미지 오브젝트를 압축하도록 구성되는 압축 유닛과; 상기 압축 유닛에 의해 압축된 이미지 오브젝트와, 상기 추출 유닛에 의해 추출된 상기 주석 오브젝트 및 상기 텍스트 오브젝트에 기초하여, 페이지 데이터를 생성하도록 구성되는 생성 유닛을 포함한다.

[0005] 본 발명의 추가적인 특징은 첨부된 도면과 관련한 예시적인 실시예에 대한 이하의 설명으로부터 명확해질 것이다.

도면의 간단한 설명

[0006] 도 1은 화상 처리 장치의 하드웨어 구성의 일례를 도시하는 도면이다.

도 2는 문서 편집 애플리케이션의 논리 구성의 일례를 나타내는 블록도이다.

도 3은 제1 실시예에 따른 문서 파일 보존 처리의 흐름을 나타내는 흐름도이다.

도 4는 문서 파일 보존의 과정을 도시하는 도면이다.

도 5는 고 압축 처리의 일례를 도시하는 도면이다.

도 6은 고 압축 처리의 일례를 도시하는 도면이다.

도 7은 이미지가 공유되고 있는 경우의 일례를 도시하는 도면이다.

도 8은 제2 실시예에 따른 문서 파일 보존 처리의 흐름을 나타내는 흐름도이다.

도 9는 제3 실시예의 과제를 설명하는 도면이다.

도 10은 제3 실시예에 따른 문서 파일 보존 처리의 흐름을 나타내는 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0007] 이하, 첨부된 도면을 참조하여, 본 발명을 바람직한 실시예에 따라 상세하게 설명한다. 이하의 실시예들에 도시된 구성은 단지 예이며 본 발명은 개략적으로 도시된 구성으로 제한되지 않는다.

[0008] (제1 실시예)

[0009] 도 1은, 본 실시예에 따른, 화상 처리 장치의 하드웨어 구성의 일례를 도시하는 도면이다. 화상 처리 장치(100)는, 예를 들어 PC 등이며, CPU(101), RAM(102), HDD(103), 범용 인터페이스(I/F)(104), 모니터(107) 및

메인 버스(108)를 구비한다. 그리고, 범용 I/F(104)에 의해, 마우스 및 키보드 같은 입력 장치(105) 및 메모리 카드 같은 외부 메모리(106)가 메인 버스(108)에 접속된다. CPU(101)는, HDD(103)에 저장된 미리결정된 애플리케이션(컴퓨터 프로그램)을 동작시킴으로써, 문장 및 표의 작성 및 편집 같은 다양한 종류의 처리를 실현한다. 이하에서, 도 1에 도시하는 구성의 화상 처리 장치(100)에서, CPU(101)가 문서 편집 애플리케이션을 실행함으로써, 문서 파일에 포함되는 주석 및 텍스트의 정보를 유지하면서, 파일 사이즈를 작게하여 문저 파일을 보존하는 처리에 대해서 설명한다.

[0010] (화상 처리 장치의 논리 구성)

[0011] 도 2는, 문서 편집 애플리케이션의 논리 구성의 일례를 나타내는 블록도이다. 임의의 애플리케이션에 의해 작성된 보존 대상의 문서 파일은 먼저 오브젝트 제어 유닛(201)에 입력된다. 여기서, 보존 대상 문서 파일로서, 이미지, 텍스트, 패스(도형), 주석 같은 다양한 속성의 오브젝트(이하, "OBJ"라고 표기)를 개별적으로 저장할 수 있는 포맷의 문서 파일을 상정하고 있다. 이 경우의 포맷으로서는, 예를 들어 전술한 PDF(Portable Document Format)를 들 수 있다.

[0012] 오브젝트 제어 유닛(201)은, 먼저, 입력 문서 파일의 각 페이지로부터 주석 및 텍스트 오브젝트를 일시적으로 추출하여 오브젝트를 저장하는 처리(추출 처리)를 행한다. 또한, 오브젝트 제어 유닛(201)은 후술하는 압축 처리가 실시된 페이지 데이터에 저장된 주석 OBJ 및 텍스트 OBJ를 복귀시키는 처리(복귀 처리)를 행한다. 또한, 오브젝트 제어 유닛(201)은 주석 및 텍스트 오브젝트가 추출된 페이지 전체를 1개의 이미지의 오브젝트로 변환하는 처리(변환 처리)도 행한다. 이들의 처리에 의해, 오브젝트 제어 유닛(201)은, 입력 문서 파일의 각 페이지에 대해서, 최초의 상태의 외관과 동일한 외관을 유지하면서, 각 페이지에 대한 파일 사이즈를 감소시켜 획득한 보존용 데이터를 생성한다. 생성된 보존용 데이터는 문서 보존 유닛(203)에 전달된다. 본 실시예에서는, 오브젝트 제어 유닛(201)이, 추출 처리, 복귀 처리, 및 변환 처리의 모두를 담당하는 구성으로 하고 있지만, 처리마다 별개의 모듈을 제공해도 된다.

[0013] 고 압축 처리 유닛(202)은, 페이지 내에 있는 이미지 OBJ(주석 및 텍스트 OBJ를 제거한 페이지 전체의 이미지 데이터를 생성하여 얻은 전체 페이지 이미지 OBJ를 포함함)의 영역을 분할하고 각 영역에 적합한 압축 처리를 행한다.

[0014] 문서 보존 유닛(203)은, 오브젝트 제어 유닛(201)으로부터 수취한 각 페이지마다 보존용 데이터를 1개의 파일에 통합하고, 그 파일을 출력 문서 파일로서 HDD(103)에 보존한다.

[0015] (문서 파일의 보존 흐름)

[0016] 상기에 이어서, 문서 편집 애플리케이션에서의 문서 파일 보존의 일련 흐름에 대해서 설명한다. 도 3은 본 실시예에 따른 문서 파일 보존 처리의 흐름을 나타내는 흐름도이다. 도 2에서 나타낸 각 처리 유닛에 의해 실현되는 일련의 처리는, CPU(101)가 HDD(103)에 저장되어 있는 미리결정된 프로그램을 RAM(102)에 로드하고 그 프로그램을 실행함으로써 실현된다.

[0017] 단계 301에서는, 오브젝트 제어 유닛(201)이 입력 문서 파일을 취득한다. 예를 들어, 오브젝트 제어 유닛(201)은 하드 디스크(103)에 보존되어 있는 문서 파일을 읽어들여서 문서 파일을 RAM(102) 위로 전개한다.

[0018] 단계 302에서는, 오브젝트 제어 유닛(201)은, RAM(102)에 전개된 문서 파일의 페이지 중, 처리 대상으로서 주목하는 페이지(이하, "주목 페이지")를 결정한다. 도 4는, 문서 파일이 보존될 때까지의 과정을, 구체예와 함께 도시하는 도면이다. 도 4에서, 주목 페이지로서의 페이지(400)는, 텍스트 OBJ(401), 주석 OBJ(402), 이미지 OBJ(403), 및 패스 OBJ 군(404)을 포함한다. 텍스트 OBJ(401)는, "비오는 날"의 2개의 문자열(색: 흑색)이 수직으로 배열된 텍스트 속성의 오브젝트이다. 이미지 OBJ(403)는, 산을 피사체로 하며, 흑색 문자의 "맑은 날" 및 적색 문자의 "산"이 수직으로 배열된 텍스트 정보(이미지내 텍스트)(410)를 포함하는 이미지 속성의 오브젝트이다. 여기서, 텍스트 OBJ(401)는, 문자 코드, 폰트 정보, 문자 피치 정보 등에 의해 묘화되는 문자 데이터(코드 정보)이다. 한편, 이미지내 텍스트(410)는 단순한 이미지 데이터(문자 코드를 포함하지 않는 화상 정보)이다. 상술한 바와 같이, 텍스트 OBJ(401)와 이미지내 텍스트(410)는 서로 성질이 상당히 다른 데이터이다. 패스 속성의 오브젝트 집합인 패스 OBJ 군(404)은 복수의 원이 겹쳐 있는 도형을 나타내고 있다. 패스 OBJ 군(404)에 포함되는 각 패스 OBJ는 복수의 점을 연결하는 직선을 포함하고, 점의 수를 증가시킴으로써 복잡한 형상을 표현할 수 있다. 그리고, 직선의 색 및 굵기의 지정, 및 직선에 의해 둘러싸이는 영역이 존재할 경우에는 그 영역 내의 색의 지정과 광이 투과될지 여부의 지정이 가능하다. 다수의 색을 포함하는 복잡한 물체를 묘화하는 경우에는, 다수의 작은 패스 OBJ를 사용한다. 도 4에 도시하는 패스 OBJ 군(404)은 10000개의 패스 OBJ를

포함하는 것으로 한다. "중요"라는 문자를 말풍선으로 나타내는 주석(402)은, 본문을 구성하는 상술한 OBJ(텍스트 OBJ(401), 이미지 OBJ(403), 패스 OBJ 군(404))과는 상이한, 본문의 보충 등을 행하는 주석 속성의 오브젝트이다. 주석 OBJ(402)의 내용은 각 OBJ와 마찬가지로 텍스트, 이미지, 패스 묘화 등으로 표현된다. 본문을 구성하는 각 OBJ에는, 그 속성을 나타내는 속성 정보가 부가되어 있어, 주석 OBJ에는, 거기에 부가하여, OBJ가 주석인 것을 나타내는 속성 정보도 부가되어 있다. 또한, 주석 OBJ의 표시 유무, 회전 방향, 위치 등을, 본문을 구성하는 각 OBJ와 별개로 그리고 독립적으로 제어할 수 있다. 도 3의 흐름 설명으로 되돌아간다.

[0019] 단계 303에서는, 오브젝트 제어 유닛(201)이, 주목 페이지 내에 주석 OBJ 또는 텍스트 OBJ가 있는지를, 각 OBJ에 부가된 속성 정보에 기초하여 판정한다. 주목 페이지 내에 주석 OBJ 또는 텍스트 OBJ가 존재하는 경우, 처리는 단계 304로 진행한다. 한편, 주목 페이지 내에 주석 OBJ 또는 텍스트 OBJ가 존재하지 않으면, 처리는 단계 305로 진행한다. 도 4의 예에서는, 페이지(400) 내에 주석 OBJ(402)와 텍스트 OBJ(401)의 양자 모두가 존재하므로, 결과적으로 처리는 단계 304로 진행된다.

[0020] 단계 304에서는, 오브젝트 제어 유닛(201)이, 주목 페이지로부터 주석 OBJ 및 텍스트 OBJ를 추출하고 OBJ를 (RAM(102)에) 저장한다. 주석 OBJ의 경우에는, 주석의 외관(묘화 내용)을 구성하는 정보와 OBJ가 주석인 것을 나타내는 정보가 저장되게 된다. 구체적으로는, 페이지 내의 표시 위치를 나타내는 위치 좌표, 색, 및 기울기 이외에, 문자 주석의 경우에는 폰트명, 폰트 사이즈, 및 문자 코드열이 저장되고, 도형의 주석의 경우에는 시점, 종점, 선 굵기, 실선 및 파선 같은 형태가 저장되며, 이미지 주석의 경우에는 그 이미지 데이터, 확대/축소 비율 등이 저장된다. 또한, 텍스트 OBJ의 경우에는, 그 문자를 표시하는 위치 좌표, 폰트명, 폰트 사이즈, 문자 코드열, 색, 및 기울기와 같은 정보가 저장되게 된다. 도 4의 예에서는, 페이지(400)로부터 주석 OBJ(402)와 텍스트 OBJ(401)의 데이터가 추출되어 저장되게 된다. 이에 의해, 페이지(400)에는, 이미지 OBJ(403)와 패스 OBJ 군(404) 만이 남는다.

[0021] 단계 305에서는, 오브젝트 제어 유닛(201)이, 주목 페이지 내에 이미지 OBJ 또는 패스 OBJ가 존재하는지의 여부를, 각 OBJ에 부가된 속성 정보에 기초하여 판정한다. 주목 페이지 내에 이미지 OBJ 또는 패스 OBJ가 존재하는 경우, 처리는 단계 306로 진행된다. 한편, 주목 페이지 내에 이미지 OBJ 또는 패스 OBJ가 존재하지 않는 경우, 처리는 단계 313로 진행된다. 도 4의 예에서는, 페이지(400) 내에 이미지 OBJ(403) 및 패스 OBJ 군(403)이 존재하므로, 처리는 단계 306로 진행되게 된다.

[0022] 단계 306에서는, 오브젝트 제어 유닛(201)이, 주목 페이지 내에 존재하는 이미지 OBJ의 수 또는 패스 OBJ의 수가 미리결정된 임계치 이상인지 여부를 판정한다. 미리결정된 임계치는 CPU(101)의 처리 능력 등을 고려하여 결정하면 된다. 이때, 이미지 OBJ와 패스 OBJ의 총 수를 구하고, 구한 총 수와 1개의 임계치를 비교하거나, 이미지 OBJ 및 패스 OBJ를 위한 임계치를 각각 설정하고, 이미지 OBJ의 수와 패스 OBJ의 수 중 어느 것이 대응하는 임계치 이상인지를 판정해도 된다. 패스 OBJ의 사이즈는 이미지 OBJ의 사이즈에 비해 작기 때문에, 파일 사이즈에 영향을 주는 정도는 작다. 그러나, 대량의 패스 OBJ가 존재하는 경우, 묘화 처리에 많은 시간이 걸리므로, 패스 OBJ의 이미지 데이터를 생성함으로써 패스 OBJ를 하나의 OBJ로 변경하는 것이 바람직하다. 미리결정된 임계치는 이러한 점을 고려하여 설정하면 된다. 판정의 결과, 이미지 OBJ의 수 또는 패스 OBJ의 수가 임계치 이상이며, 처리는 단계 307로 진행한다. 한편, 이미지 OBJ의 수 및 패스 OBJ의 수가 임계치 미만이면, 처리는 단계 308로 진행한다. 임계치를 1000개로 하면, 도 4의 예에서는, 페이지(400) 내의 이미지 OBJ의 수는 1개이지만, 패스 OBJ의 수는 10000개이므로, 처리는 단계 307로 진행하게 된다.

[0023] 단계 307에서는, 오브젝트 제어 유닛(201)이 주목 페이지 전체의 이미지 데이터를 생성한다. 이 경우의 주목 페이지이란, 주석 OBJ와 텍스트 OBJ가 추출된 페이지(이미지 OBJ와 패스 OBJ만을 포함하는 페이지)이다. 페이지 전체의 이미지 데이터의 생성에 의해, 페이지 전체는 1개의 이미지 OBJ가 된다. 도 4에서의 부호(420)는, 주석 OBJ와 텍스트 OBJ가 추출된 페이지 전체(400)의 이미지 데이터를 생성함으로써 얻어진 페이지 전체의 이미지 OBJ를 나타낸다. 이 이미지 OBJ(420)는 이미지 OBJ(403)와 패스 OBJ 군(404)의 묘화 내용을 포함한다. 상기와 같이 일정수 이상의 이미지 OBJ와 패스 OBJ를 포함하는 페이지 전체를 1개의 이미지 OBJ로 변환함으로써, 후속 묘화 처리 및 인쇄 처리에 필요한 시간을 단축할 수 있다. 페이지 전체의 이미지 데이터를 생성한 후에, 처리는 단계 308로 진행한다.

[0024] 단계 308에서는, 오브젝트 제어 유닛(201)이 주목 페이지의 이미지 OBJ로부터 이미지 OBJ를 결정한다. 이 경우, 주목 페이지의 이미지 OBJ는, 주목 페이지 내에 원래 존재하는 이미지 OBJ 외에, 단계 307에서 페이지 전체의 이미지 데이터를 생성함으로써 얻어진 페이지 전체의 이미지 OBJ를 포함한다. 결정된 주목 이미지 OBJ의 데이터는 고 압축 처리 유닛(202)에 전달된다.

[0025] 단계 309에서는, 고 압축 처리 유닛(202)이, 주목 이미지 OBJ를 텍스트 영역과 텍스트 영역 이외의 배경 영역으로 분할하고, 각각의 영역에 적합한 방식으로 각각의 영역을 압축한다. 주목 이미지 OBJ 내에 문자 정보가 없으면, 본 단계는 생략되고, 처리는 다음 단계 310로 진행한다. 이하, 본 단계에서의 고 압축 처리에 대해서 설명한다.

[0026] 먼저, 주목 이미지 OBJ는 2치화되고, 얻어진 모노크롬 2치 화상에 대하여 영역 분할 처리를 행한다. 이 경우의 영역 분할은, 예를 들어 일본 특허 공개 공보 제2000-332984호에 기재된 바와 같은 일반적인 레이아웃 분석 기술일 수 있다. 도 5는, 전술한 구체예의 케이스(도 4)에서의 고 압축 처리의 내용을 도시하는 도면이다. 페이지 전체의 이미지 데이터가 생성된 이미지 OBJ(420)의 경우, 이미지내 텍스트(410)의 부분이 텍스트 영역(501)이 되고 다른 부분이 배경 영역(502)이 되도록 영역 분할이 행해진다. 그리고, 텍스트 영역(501)에 대하여는, "맑은 날" 및 "산"을 나타내는 2치 화상(문자 화상(510))이 생성된다. 또한, 텍스트 영역(501)에 대해서는, 이미지 OBJ(420)와 문자 화상(510)으로부터 문자색이 추출되며, 텍스트 영역(501)에 대응하는 문자색 화상(520)이 생성된다. 이 문자색 화상(520)은, "맑은 날"의 문자색을 나타내는 흑색 직사각형(521)과 "산"의 문자색을 나타내는 적색 직사각형(522)을 포함하는 다치의 이미지 속성의 화상이다. 그리고, 배경 영역(502)에 대해서는, 문자 화상(510)에서의 문자 영역 부분에 대해 채우기 처리를 행함으로써 얻어지는 다치 배경 화상(530)이 생성된다. 그리고, 2치 문자 화상(510)에 대해서는, 가역 방식의 압축 처리(예를 들어, MMR 압축)가 행해지며, 다치 문자색 화상(520) 및 다치 배경 화상(530)에 대해서는, 비가역 방식의 압축 처리(예를 들어, JPEG 압축)가 각각 행해진다. 텍스트 영역의 도형문자 정보에 대해서 가역 압축을 적용하는 이유는 압축해제 시의 화상 품질이 열화되는 것을 억제하여 가독성을 유지하기 위해서이다. 이상과 같이 하여, 문자 정보를 포함하는 이미지 OBJ는, 3종류의 압축 화상, 즉 문자 화상(2치), 문자색 화상(다치), 및 배경 화상(다치)으로 변환된다.

[0027] 단계 310에서는, 오브젝트 제어 유닛(201)이 주목 페이지 내에 처리되지 않은 이미지 OBJ가 있는지를 판정한다. 판정의 결과, 주목 페이지 내에 처리되지 않은 이미지 OBJ가 있으면, 처리는 단계 308로 복귀되고, 다음 이미지 OBJ를 결정하고 처리를 계속한다. 한편, 주목 페이지 내의 이미지 OBJ가 모두 이미 처리되어 있는 경우, 처리는 단계 311로 진행한다. 단계 307에서 페이지 전체의 이미지 데이터가 생성된 경우, 페이지 내에는 처리되지 않은 이미지 OBJ가 남아있지 않으므로, 처리는 즉시 단계 311로 진행하게 된다.

[0028] 단계 311에서는, 오브젝트 제어 유닛(201)은, 주목 페이지로부터 추출되어 저장된 주석 OBJ 또는 텍스트 OBJ가 있는지를 확인한다. RAM(102)에 주석 OBJ 또는 텍스트 OBJ의 데이터가 저장되어 있는 경우, 처리는 단계 312로 진행한다. 한편, RAM(102)에 주석 OBJ 또는 텍스트 OBJ의 데이터가 저장되어 있지 않은 경우, 처리는 단계 313로 진행한다.

[0029] 단계 312에서는, 오브젝트 제어 유닛(201)은, 저장된 주석 OBJ와 텍스트 OBJ를, 그 외관이 유지되도록 주목 페이지로 복귀시키고, 출력 파일로서 새로운 페이지 데이터를 생성한다. 즉, 주석 OBJ와 텍스트 OBJ는, 그 외관이 위치, 크기, 회전각 등에 대해 원래의 페이지와 완전히 동일하도록, 주목 페이지로 복귀된다. 상술한 바와 같이, 본 실시예에서는, 이미지 OBJ는, 총 3종류의 이미지, 즉 텍스트 영역을 구성하는 2개의 화상(문자 화상과 문자색 화상)과 텍스트 영역 이외의 영역을 구성하는 배경 화상으로 분할된다. 새로운 페이지 데이터에서는, 페이지는 배경 화상이 문자 화상과 문자색 화상 하에 페이지 내의 동일 위치 및 동일 사이즈로 묘화되도록 페이지가 기술된다. 주석 OBJ가 이미지 OBJ에 겹치는 경우, 페이지는 주석 OBJ가 3개의 이미지 위에 또한 묘화되도록 기술된다.

[0030] 단계 313에서는, 오브젝트 제어 유닛(201)이, RAM(102)에 전개된 문서 파일의 모든 페이지를 처리했는지를 판정한다. 처리되지 않은 페이지가 있는 경우, 처리는 단계 302로 복귀되고, 다음 주목 페이지를 결정하고 처리를 계속한다. 한편, 모든 페이지가 처리되었으면, 처리는 단계 314로 진행한다.

[0031] 단계 314에서는, 문서 보존 유닛(203)이, 각 페이지 데이터를 통합해서 1개의 보존용 문서 파일을 생성하고, 문서를 HDD(103)에 보존한다.

[0032] 이상이, 본 실시예에 따른 문서 파일 보존 처리의 내용이다.

[0033] 상술한 고 압축 처리에서는, 영역 분할에 의해 얻어진 텍스트 영역이 2종류의 압축된 이미지, 즉 문자 화상(2치)과 문자색 화상(다치)으로 변환되지만, 문자색 정보를 이미지가 아닌 문자 코드로 나타낼 수도 있다. 도 6은, 텍스트 영역에서의 문자색 정보를 문자 코드로 나타낼 경우의 고 압축 처리의 일례를 도시하는 도면이다. 이 예에서는, 이미지 OBJ(420)에 대하여 영역 분할을 행하는 것 이외에, 문자 부분의 색 판정을 행한다. 그리고, 문자 부분의 색 판정 결과에 기초하여, 문자색마다 2치 문자 화상을 생성한다. 도 6의 예에서는, 흑색

문자 "맑은 날"에 대응하는 문자 화상(601) 및 적색 문자 "산"에 대응하는 문자 화상(602)의 2개의 문자 화상이 생성된다. 그리고, 이들 2개의 문자 화상(2치)에 대하여, 가역 방식의 압축 처리가 행해진다. 그리고, 색 판정 결과를 따른 문자색 코드(흑색과 적색 문자색 코드)가, 2개의 문자 화상(601 및 602)의 압축된 화상과 각각 연관지어진다. 이와 같이, 영역 분할에 의해 얻어진 텍스트 영역을 문자색마다 문자 화상(2치)과 문자색 코드로 변환하고, 문자색마다 생성한 문자 화상에 압축 처리를 행해도 된다.

[0034] 또한, 이미지 OBJ를 영역 분할에 의해 텍스트 영역과 배경 영역으로 분리하여 압축을 행하는 대신에, 이미지의 해상도를 감소시키거나 압축 방식을 더 높은 압축율을 갖는 방식으로 변경함으로써 이미지 OBJ의 데이터 사이즈를 감소시켜도 된다.

[0035] 본 실시예에 따르면, 보존 대상 문서 파일의 페이지 내에 주석이나 텍스트가 포함되어 있는 경우에, 정보를 유지하면서 파일 사이즈를 감소시킬 수 있다. 또한, 미리결정된 조건 하에서 페이지 전체의 이미지 데이터가 생성된 경우에는, OBJ의 수가 감소되므로, 보존된 문서 파일에서는 페이지 묘화 처리에 필요한 시간을 짧게할 수 있다.

[0036] (제2 실시예)

[0037] 제1 실시예는, 주석 및 텍스트가 추출된 페이지 내의 이미지 OBJ의 수 또는 패스 OBJ의 수가 임계치 이상인 경우에, 페이지 전체의 이미지 데이터를 생성하는 양태이다. 그러나, 복수의 페이지에 의해 이미지 OBJ의 리소스가 공유되는 경우, 제1 실시예의 기술에서는 도리어 파일 사이즈가 증가할 가능성이 있다. 여기서, 이미지 OBJ의 리소스가 공유되는 일례를 설명한다. 도 7은, 총 3 페이지, 즉 페이지(710 내지 730)를 포함하는 문서 파일에서 이미지가 공유되는 예이다. 페이지(710) 내에는 이미지 OBJ(711)가 포함되고, 페이지(720) 내에는 이미지 OBJ(721)가 포함되며, 페이지(730) 내에는 이미지 OBJ(731)가 포함된다. 그리고, 이것들 3개의 이미지 OBJ(711, 721, 731)는 1개의 이미지 리소스(700)를 공유하고 있다. 또한, 각 페이지에 포함되는 패스 OBJ 군(712, 722, 723)의 각각은 10000개의 패스 OBJ로 구성되며, 각 페이지에서 독립적으로 그 데이터가 유지된다. 여기서, 가령 이미지 리소스(700)의 데이터 사이즈가 100MB인 것으로 하고, 3개의 패스 OBJ 군의 데이터 사이즈가 10MB인 것으로 하면, 문서 전체의 파일 사이즈는 110MB이다. 제1 실시예에서, 이미지 OBJ의 수 또는 패스 OBJ의 수가 임계치 이상인 경우, 각 페이지 전체의 이미지 데이터가 생성되며, 각 페이지 전체의 이미지 데이터의 생성 결과로 이미지 영역이 커지기 때문에 각 페이지의 데이터 사이즈가 150MB가 되는 것으로 한다. 이 경우, 각 페이지에 존재하는 이미지 OBJ의 데이터 사이즈가 압축 처리에 의해 50MB가 되더라도, 문서 전체의 파일 사이즈는 50MB×3(페이지)=150MB이 된다. 이는 원래 파일 사이즈인 110MB보다 크다.

[0038] 결과적으로, 페이지 내의 이미지 OBJ의 수 또는 패스 OBJ의 수가 임계치 이상인 경우에도, 복수의 페이지에 의해 공유되는 이미지 OBJ가 존재하는 조건에서는 페이지 전체의 이미지 데이터가 생성되지 않는 양태가 제2 실시예로서 설명된다. 제1 실시예의 것과 공통되는 내용에 대한 설명은 생략하거나 간략화하고, 이하에서는 상이한 점만을 설명한다.

[0039] 도 8은 본 실시예에 따른 문서 파일 보존 처리의 흐름을 도시하는 흐름도이다. 단계 801 내지 806은 제1 실시예의 도 3의 흐름에서의 단계 301 내지 306에 대응한다. 즉, 취득된 보존 대상의 문서 파일 내의 주목 페이지에 주석 OBJ 또는 텍스트 OBJ가 있는 경우, OBJ는 페이지로부터 추출되고 저장된다(S801 내지 S804). 또한, 주목 페이지 내에 이미지 OBJ 또는 패스 OBJ가 있는 경우, 이미지 OBJ와 패스 OBJ의 총 수가 임계치 이상인지 여부가 판정된다(S805, S806). 판정의 결과, 이미지 OBJ 및 패스 OBJ의 수가 임계치 이상인 경우, 처리는 단계 807로 진행한다. 한편, 이미지 OBJ 및 패스 OBJ의 수가 임계치 미만인 경우, 처리는 단계 810로 진행한다.

[0040] 단계 807에서는, 복수의 페이지에 의해 리소스가 공유되는 이미지 OBJ(이하, 공유 이미지 OBJ)가 주목 페이지 내에 존재하는 지의 여부가 판정된다. 공유 이미지 OBJ가 존재하는 경우는, 처리는 단계 808로 진행한다. 한편, 공유 이미지 OBJ가 존재하지 않는 경우에는, 처리는 단계 809로 진행한다.

[0041] 단계 808에서는, 공유 이미지 OBJ를 제외한 페이지 전체의 이미지 데이터가 생성된다. 즉, 주석 OBJ 및 텍스트 OBJ 이외에 공유 이미지 OBJ가 제거된 페이지(기타의 이미지 OBJ와 패스 OBJ 만을 포함하는 페이지)가 본 단계에서의 페이지 전체의 이미지 데이터를 생성하는 처리의 대상이 된다. 페이지 전체의 이미지 데이터를 생성하는 처리에 의해, 공유 이미지 OBJ를 제외한 페이지 전체가 하나의 이미지 OBJ가 된다. 페이지 전체의 이미지 데이터를 생성하는 처리 후에, 처리는 주목 이미지 OBJ가 결정되는 단계 810으로 진행한다.

[0042] 단계 809에서는, 도 3의 흐름에서의 단계 307과 마찬가지로, 주석 OBJ 및 텍스트 OBJ가 제거된 페이지 전체의 이미지 데이터가 생성된다. 페이지 전체의 이미지 데이터를 생성한 후에, 처리는 주목 이미지 OBJ가 결정되는

단계 810으로 진행한다.

[0043] 단계 810에서는, 주목 페이지 내의 이미지 OBJ로부터 주목 이미지 OBJ가 결정된다. 그리고, 주목 이미지 OBJ에 대하여, 전술한 고 압축 처리가 실시된다(단계 811, 도 3의 흐름의 단계 309에 대응). 후속 단계 812 내지 816의 각각에서의 처리는, 도 3의 흐름에서의 단계 310 내지 314 각각의 처리에 대응하며, 그들 사이에 특히 상이한 점을 없기 때문에, 설명을 생략한다.

[0044] 이상과 같은 처리에 의해, 예를 들어 전술한 도 7의 예에서는, 이미지 리소스의 데이터 사이즈가 50MB, 3개의 패스 OBJ 군의 데이터 사이즈가 10MB, 즉 총 60MB이므로, 파일 사이즈를 원래의 파일 사이즈인 110MB보다 작은 사이즈로 감소시킬 수 있다.

[0045] 본 실시예에 따라서, 복수의 페이지에 의해 이미지 리소스가 공유되는 문서 파일에서도 제1 실시예와 마찬가지의 효과를 얻을 수 있다.

[0046] (제3 실시예)

[0047] 제1 실시예에서는, 주석 OBJ 및 텍스트 OBJ를 일시적으로 삭제하게 되지만, 이에 의해 기밀 유지 같은 정보 보안의 관점에서 문제가 발생하는 경우가 있다. 도 9는 그 구체예를 나타낸다. 도 9의 예는, 청구서로서의 페이지(900) 내에 텍스트 OBJ(901)가 존재하고, 당해 텍스트 OBJ(901)에서의 수신처 부분을 불투명한 채움 직사각형 주석 OBJ(902)에 의해 시각적으로 인지되는 것을 방지하는 경우를 나타낸다. 이러한 경우에 제1 실시예를 적용하면, 주석 OBJ를 페이지로부터 제거한 후의 도중의 처리에서, 본래는 시각적으로 인지되지 않아야 하는 수신처가 보이게 될 것이다. 최종적으로, 주석 OBJ는 그 원래의 위치로 복귀되지만, 예를 들어 이 문서 포맷이 PDF인 경우, PDF를 편집할 수 있는 소프트웨어에 의해, 용이하게 당해 주석 OBJ를 이동하여 수신처를 시각적으로 인지하게 될 수 있다. 이러한 문제를 다룰 수 있는 양태를 제3 실시예로서 설명한다. 제1 실시예의 것과 공통되는 내용에 대한 설명은 생략하거나 간략화하고, 이하에서는 상이한 점만을 설명한다.

[0048] 도 10은, 본 실시예에 따른, 문서 파일 보존 처리의 흐름을 나타내는 흐름도이다. 단계 1001 내지 1003은, 제1 실시예의 도 3의 흐름에서의 단계 301 내지 303에 대응한다. 즉, 보존 대상 문서 파일이 취득되고(S1001), 주목 페이지가 결정되면(S1002), 당해 주목 페이지 내에 주석 OBJ 또는 텍스트 OBJ가 있는지 여부가 판정된다(S1003). 판정의 결과, 주목 페이지 내에 주석 OBJ 또는 텍스트 OBJ가 존재하는 경우, 처리는 단계 1004로 진행한다. 한편, 주목 페이지 내에 주석 OBJ 또는 텍스트 OBJ가 존재하지 않는 경우, 처리는 단계 1010로 진행한다.

[0049] 단계 1004에서는, 오브젝트 제어 유닛(201)이, 주목 페이지 내에서 주목하는 주석 OBJ 또는 텍스트 OBJ(이하, 주목 주석 OBJ 또는 주목 텍스트 OBJ)를 결정한다.

[0050] 단계 1005에서는, 오브젝트 제어 유닛(201)이, 주목 주석 OBJ 혹은 주목 텍스트 OBJ 가 다른 OBJ와 겹치는지 여부를 판정한다. 주목 주석 OBJ 혹은 주목 텍스트 OBJ가 다른 OBJ와 겹치는 경우, 처리는 단계 1006로 진행한다. 한편, 주목 주석 OBJ 혹은 주석 텍스트 OBJ가 다른 OBJ와 겹치는 경우, 처리는 단계 1008로 진행한다. 도 9의 예에서는, 주석 OBJ(902)와 텍스트 OBJ(901)가 서로 겹쳐 있다. 결과적으로, 그 중 하나가 주목 OBJ인 경우에도, 처리는 단계 1006로 진행하게 된다.

[0051] 단계 1006에서는, 오브젝트 제어 유닛(201)이, 서로 겹쳐 있는 OBJ 중, 다른 OBJ 위에 있는(상층측에 있는) 것이 채우기되도록 지정되어 있는지의 여부를 판정한다. 예를 들어, 상위 OBJ에 대한 색 정보에 대해 불투명한 색이 지정되어 있는 경우, 채우기가 지정되어 있다고 결정된다. 판정의 결과, 상위 OBJ가 채우기되도록 지정되어 있는 경우, 처리는 단계 1007로 진행한다. 한편, 상위 OBJ가 채우기되도록 지정되어 있지 않은 경우, 처리는 단계 1008로 진행한다. 도 9의 예에서, 주석 OBJ(902)는, 세장형 직사각형의 내부가 불투명한 흑색으로 채우기되도록 지정되어 있는 주석 OBJ이므로, 처리는 단계 1007로 진행하게 된다.

[0052] 단계 1007에서는, 오브젝트 제어 유닛(201)이, 서로 겹쳐 있는 오브젝트의 이미지 데이터를 함께 생성하고 이미지 OBJ를 생성한다. 도 9의 예에서는, 텍스트 OBJ(901)와 주석 OBJ(902)를 함께 조합하는 하나의 이미지 OBJ가 생성된다. 즉, 주석 OBJ(902)가 텍스트 OBJ(901)와 일체화된 1개의 이미지 OBJ가 되므로, 주석 OBJ(902)는 더 이상 독립적인 주석 OBJ가 아니다. 이에 의해, 후속 처리 및 편집에서, 본래는 주석에 의해 숨겨져 있는 수신처가 시각적으로 인지되는 것을 방지할 수 있다.

[0053] 단계 1008은, 도 3의 흐름에서의 단계 304에 대응하고, 주목 주석 OBJ 혹은 주목 텍스트 OBJ는 주목 페이지로부터 추출되고 저장된다.

[0054] 단계 1009에서는, 주목 페이지 내에 처리되지 않은 주석 OBJ 또는 처리되지 않은 텍스트 OBJ가 있는지 여부가 판정된다. 판정의 결과, 주목 페이지 내에 처리되지 않은 주석 OBJ 또는 처리되지 않은 텍스트 OBJ가 있는 경우, 처리는 단계 1004로 복귀되고, 다음 주석 OBJ 또는 텍스트 OBJ을 결정하며 처리를 계속한다. 한편, 주목 페이지 내의 주석 OBJ 모두 또는 텍스트 OBJ 모두가 이미 처리된 경우, 처리는 단계 1010로 진행한다.

[0055] 단계 1010 내지 1019 각각의 처리는 단계 305 내지 314 각각의 처리에 대응하며, 그들 사이에 특별히 상이한 점은 없기 때문에, 설명을 생략한다.

[0056] 전술한 도 9의 예에서는, 주석 OBJ(902)의 데이터가 단계 1008에서 저장되지 않으므로, 주석 OBJ(902)가 단계 1017에서 페이지로 복귀되지 않는다. 즉, 텍스트 OBJ(901) 내의 수신처가 주석 OBJ(902)로 채워진 상태의 페이지가 단계 1019에서 보존되게 된다.

[0057] 이상이, 본 실시예에 따른 문서 보존 처리의 내용이다.

[0058] 본 실시예에 따르면, 주석 등의 오브젝트가, 페이지 내의 기밀 정보가 시각적으로 인지되는 것을 방지하기 위해서 사용되는 경우, 기밀 정보를 유지하면서 제1 실시예의 것과 동일한 효과를 얻을 수 있다.

[0059] (다른 실시예)

[0060] 본 발명의 실시예(들)는, 전술한 실시예(들) 중 하나 이상의 기능을 실행하기 위해 저장 매체(보다 완전하게는 '비일시적 컴퓨터 판독가능 저장 매체'라 칭할수도 있음)에 기록된 컴퓨터 실행가능 명령어(예를 들어, 하나 이상의 프로그램)를 판독 및 실행하고 그리고/또는 전술한 실시예(들) 중 하나 이상의 기능을 실행하는 하나 이상의 회로(예를 들어, 주문형 집적 회로(ASIC))를 포함하는 시스템 또는 장치의 컴퓨터에 의해, 그리고 예를 들어 전술한 실시예(들) 중 하나 이상의 기능을 실행하기 위해 저장 매체로부터 컴퓨터 실행가능 명령어를 판독 및 실행함으로써 그리고/또는 전술한 실시예(들) 중 하나 이상의 기능을 실행하기 위해 하나 이상의 회로를 제어함으로써 상기 시스템 또는 장치의 컴퓨터에 의해 실행되는 방법에 의해 실현될 수도 있다. 컴퓨터는 하나 이상의 프로세서(예를 들어, 중앙 처리 유닛(CPU), 마이크로 처리 유닛(MPU))를 포함할 수 있고 컴퓨터 실행가능 명령어를 판독 및 실행하기 위한 별도의 컴퓨터 또는 별도의 프로세서의 네트워크를 포함할 수 있다. 컴퓨터 실행가능 명령어는 예를 들어 네트워크 또는 저장 매체로부터 컴퓨터에 제공될 수 있다. 저장 매체는, 예를 들어 하드 디스크, 랜덤 액세스 메모리(RAM), 리드 온리 메모리(ROM), 분산형 컴퓨팅 시스템의 스토리지, 광디스크(예를 들어, 콤팩트 디스크(CD), 디지털 다기능 디스크(DVD) 또는 블루레이 디스크(BD)™), 플래시 메모리 디바이스, 메모리 카드 등 중 하나 이상을 포함할 수 있다.

[0061] (기타의 실시예)

[0062] 본 발명은, 상기의 실시형태의 1개 이상의 기능을 실현하는 프로그램을, 네트워크 또는 기억 매체를 개입하여 시스템 혹은 장치에 공급하고, 그 시스템 혹은 장치의 컴퓨터에 있어서 1개 이상의 프로세서가 프로그램을 읽어 실행하는 처리에서도 실현가능하다.

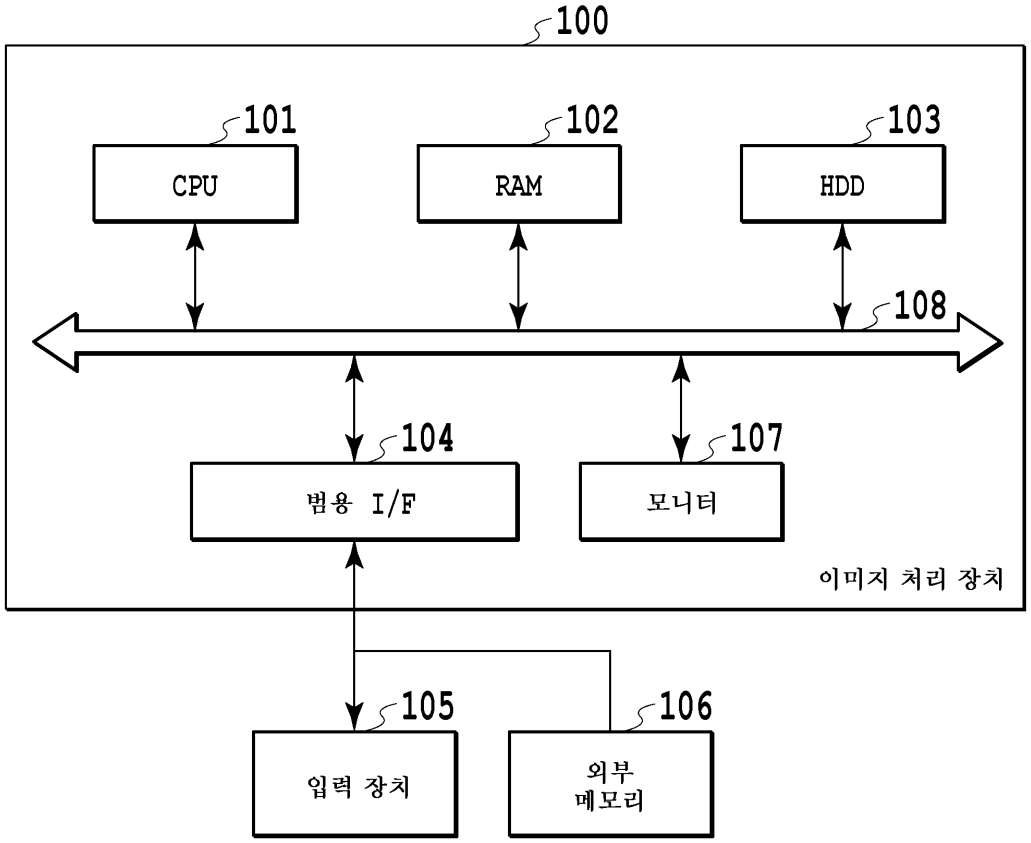
[0063] 또한, 1개 이상의 기능을 실현하는 회로(예를 들어, ASIC)에 의해서도 실행가능하다.

[0064] 본 발명에 따르면, 페이지 내에 주석 및 텍스트를 포함하는 문서 파일의 경우에도, 주석 또는 텍스트 정보를 유지하면서 파일 사이즈를 감소시킬 수 있고 문서 파일을 보존할 수 있다.

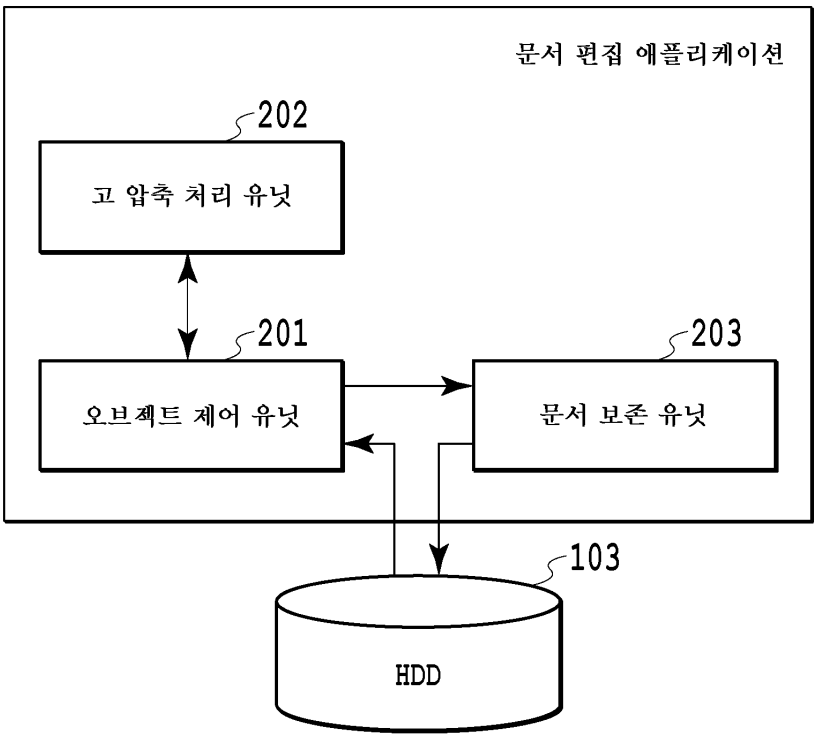
[0065] 본 발명을 예시적인 실시예와 관련하여 설명하였지만, 본 발명은 개시된 예시적인 실시예로 제한되지 않는다는 것이 이해될 것이다. 이하의 청구항의 범위는 이러한 모든 변형과 동등한 구조 및 기능을 포함하도록 최광의로 해석되어야 한다.

도면

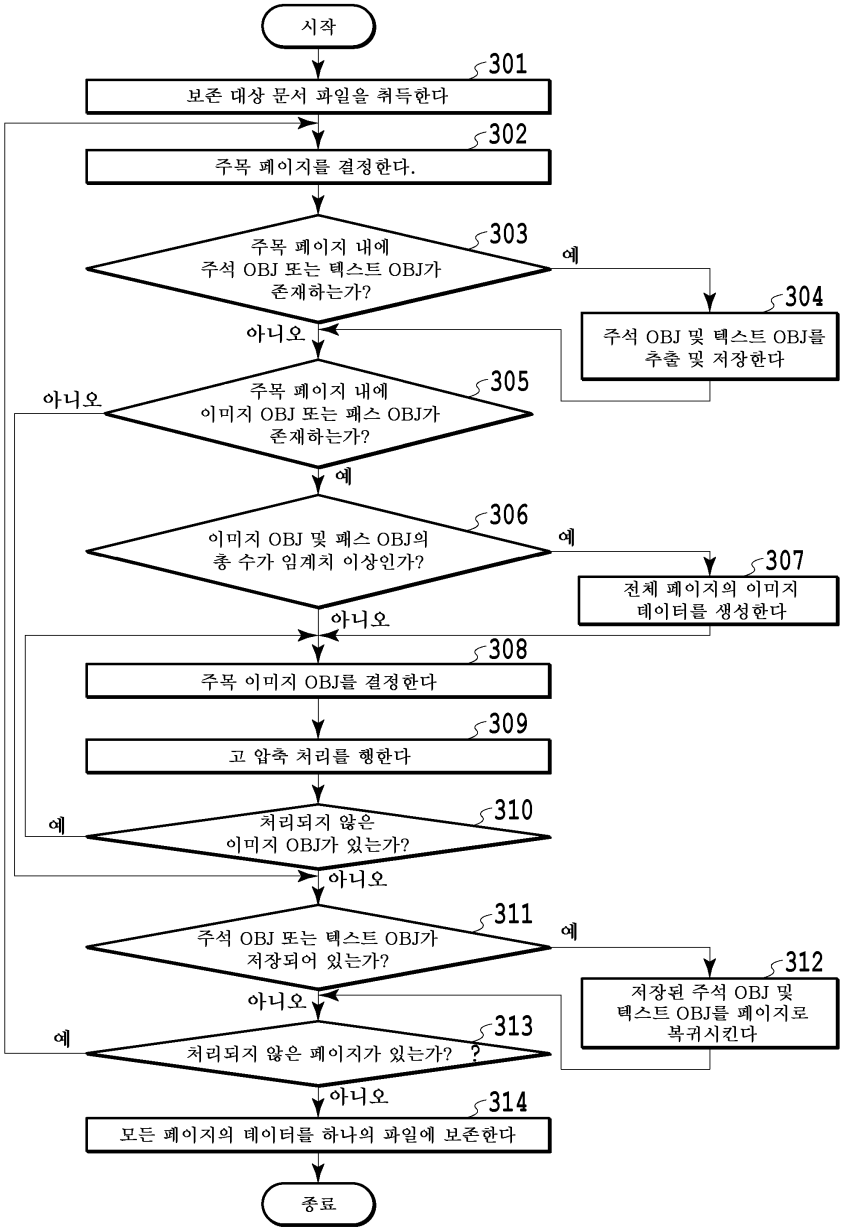
도면1


100
101
CPU
102
RAM
103
HDD
108
104
범용 I/F
107
모니터
이미지 처리 장치
105
입력 장치
106
외부
메모리

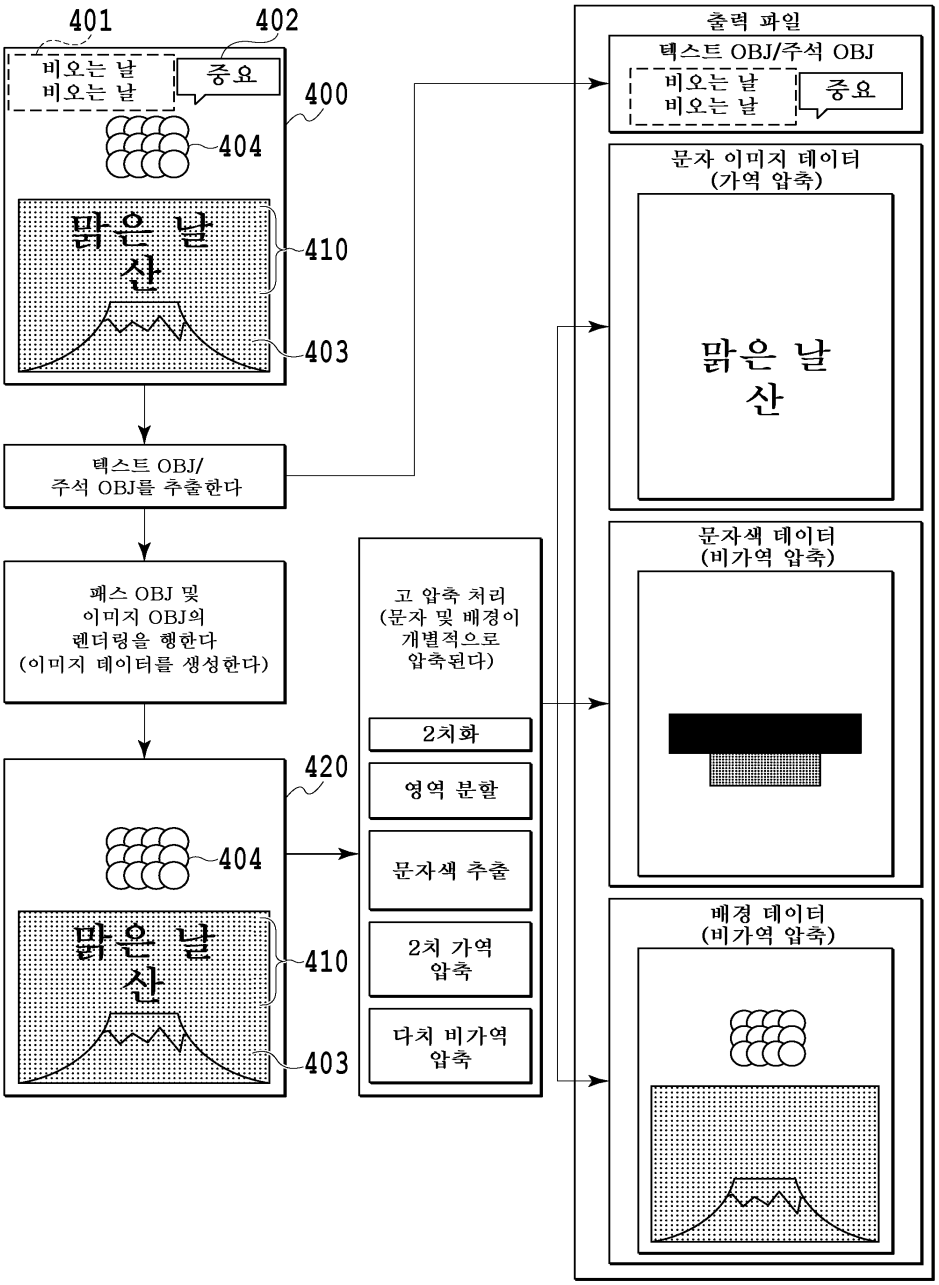
도면2


문서 편집 애플리케이션
202
고 압축 처리 유닛
201
오브젝트 제어 유닛
203
문서 보존 유닛
103
HDD

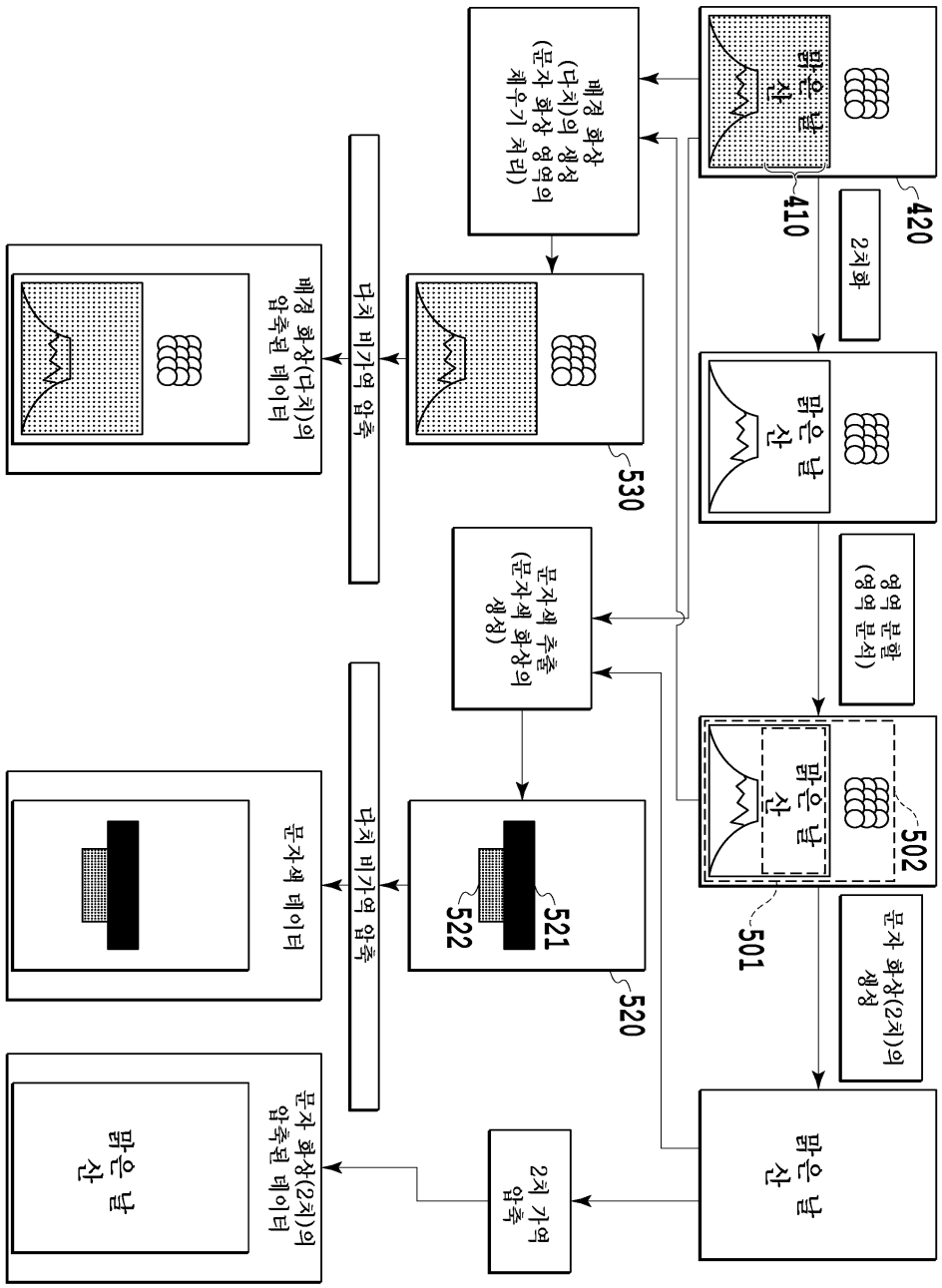
도면3


시작
301
보존 대상 문서 파일을 취득한다
302
주목 페이지를 결정한다.
303
주목 페이지 내에 주석 OBJ 또는 텍스트 OBJ가 존재하는가?
예
304
주석 OBJ 및 텍스트 OBJ를 추출 및 저장한다
아니오
305
주목 페이지 내에 이미지 OBJ 또는 패스 OBJ가 존재하는가?
아니오
예
306
이미지 OBJ 및 패스 OBJ의 총 수가 임계치 이상인가?
예
307
전체 페이지의 이미지 데이터를 생성한다
아니오
308
주목 이미지 OBJ를 결정한다
309
고 압축 처리를 행한다
310
처리되지 않은 이미지 OBJ가 있는가?
예
아니오
311
주석 OBJ 또는 텍스트 OBJ가 저장되어 있는가?
예
312
저장된 주석 OBJ 및 텍스트 OBJ를 페이지로 복귀시킨다
아니오
313
처리되지 않은 페이지가 있는가?
예
아니오
314
모든 페이지의 데이터를 하나의 파일에 보존한다
종료

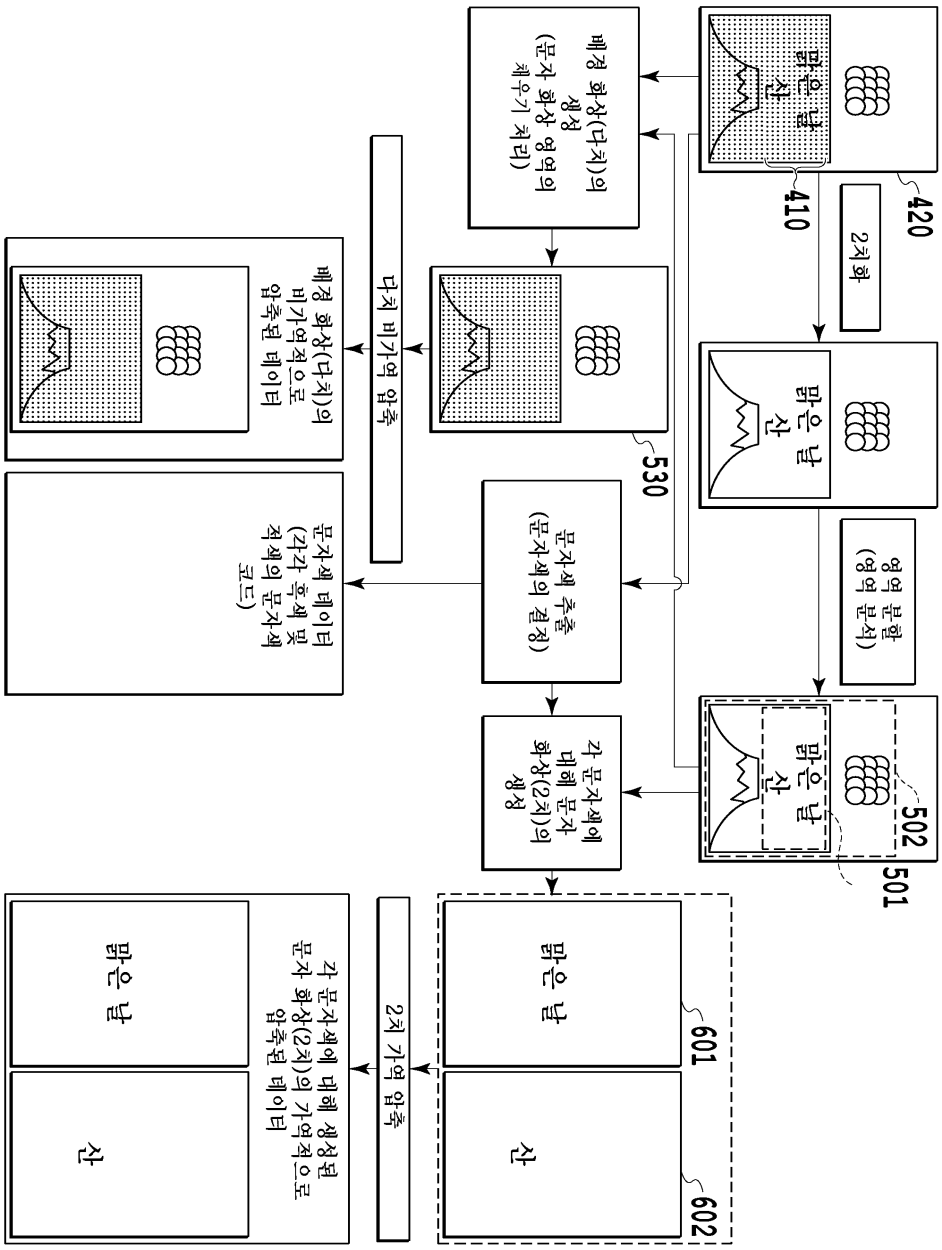
도면4


401
402
비오는 날
비오는 날
중요
400
404
맑은 날
산
410
403
텍스트 OBJ/
주석 OBJ를 추출한다
패스 OBJ 및
이미지 OBJ의
렌더링을 행한다
(이미지 데이터를 생성한다)
420
404
맑은 날
산
410
403
고 압축 처리
(문자 및 배경이
개별적으로
압축된다)
2치화
영역 분할
문자색 추출
2치 가역
압축
다치 비가역
압축
출력 파일
텍스트 OBJ/주석 OBJ
비오는 날
비오는 날
중요
문자 이미지 데이터
(가역 압축)
맑은 날
산
문자색 데이터
(비가역 압축)
배경 데이터
(비가역 압축)

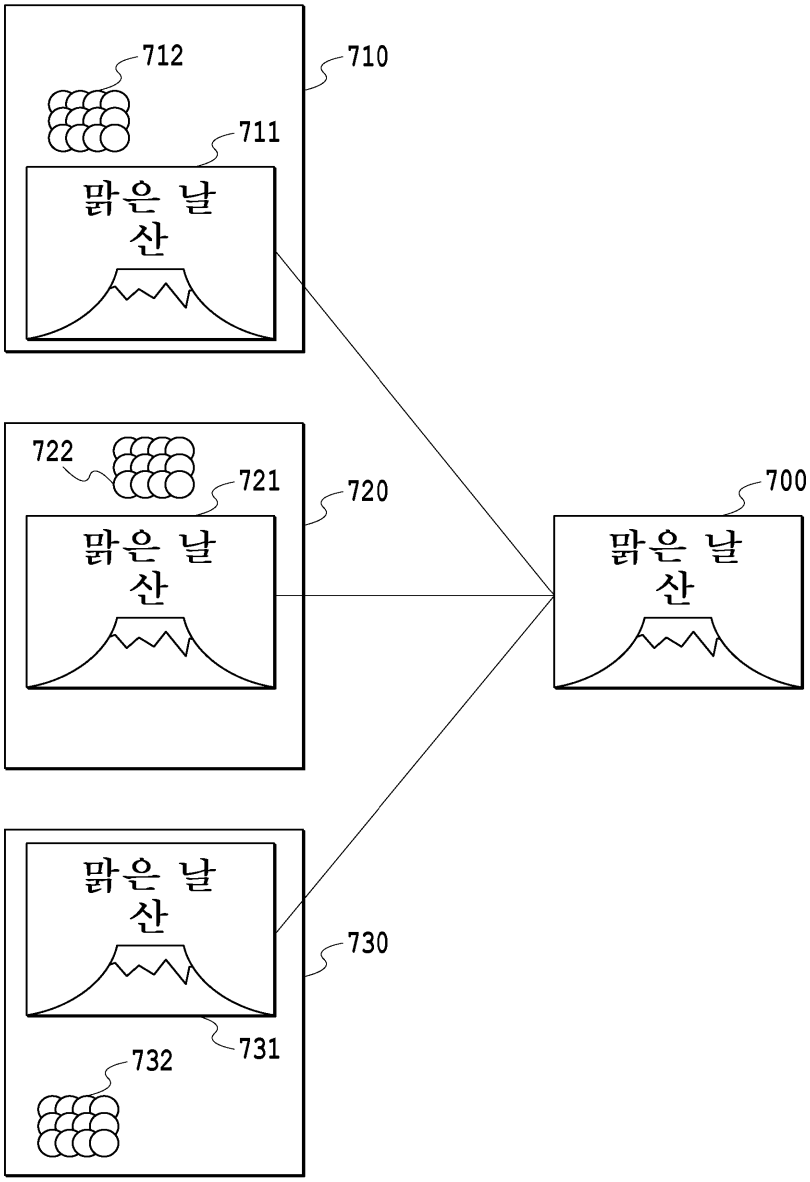
도면5


맑은 날
산
410
420
2치화
맑은 날
산
영역 분할
(영역 분석)
502
맑은 날
산
501
문자 화상(2치)의
생성
맑은 날
산
2치 가역
압축
문자 화상(2치)의
압축된 데이터
맑은 날
산
다치 비가역 압축
문자색 데이터
문자색 추출
(문자색 화상의
생성)
521
522
520
다치 비가역 압축
배경 화상
(다치)의 생성
(문자 화상 영역의
체우기 처리)
530
다치 비가역 압축
배경 화상(다치)의
압축된 데이터

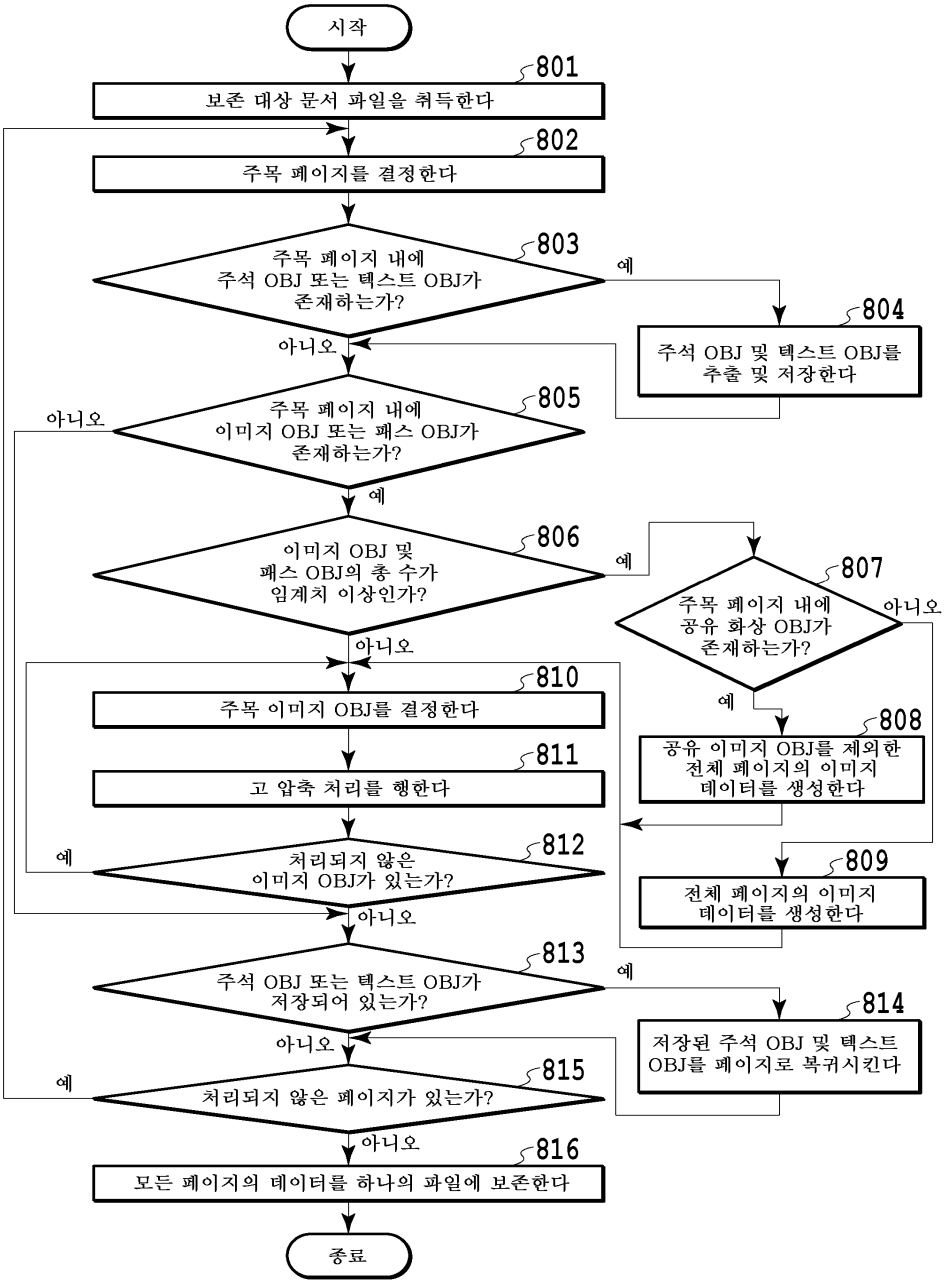
도면6


맑은 날
산
410
420
2치화
배경 화상(다치)의 생성
(문자 화상 영역의 채우기 처리)
맑은 날
산
영역 분할
(영역 분석)
530
다치 비가역 압축
배경 화상(다치)의 비가역적으로 압축된 데이터
문자색 추출
(문자색의 결정)
맑은 날
산
502
501
문자색 데이터
(각각 흑색 및 적색의 문자색 코드)
각 문자색에 대해 문자 화상(2치)의 생성
맑은 날
산
601
602
2치 가역 압축
각 문자색에 대해 생성된 문자 화상(2치)의 가역적으로 압축된 데이터
맑은 날
산

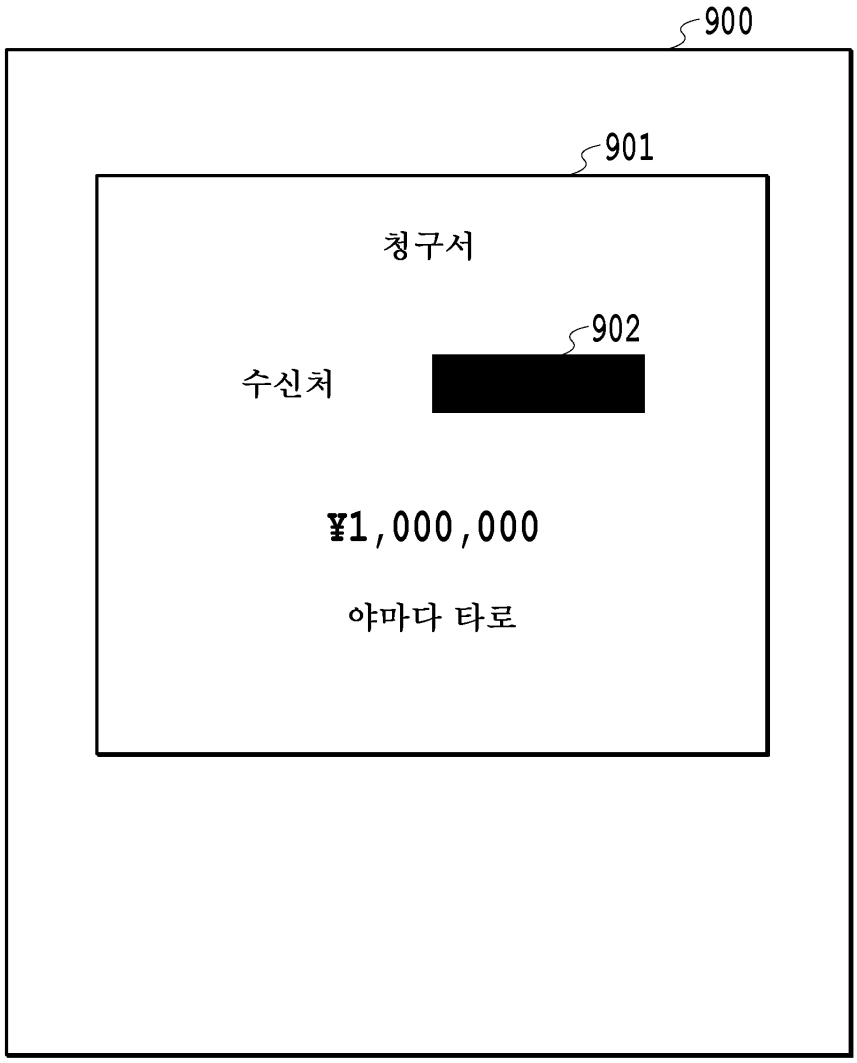
도면7



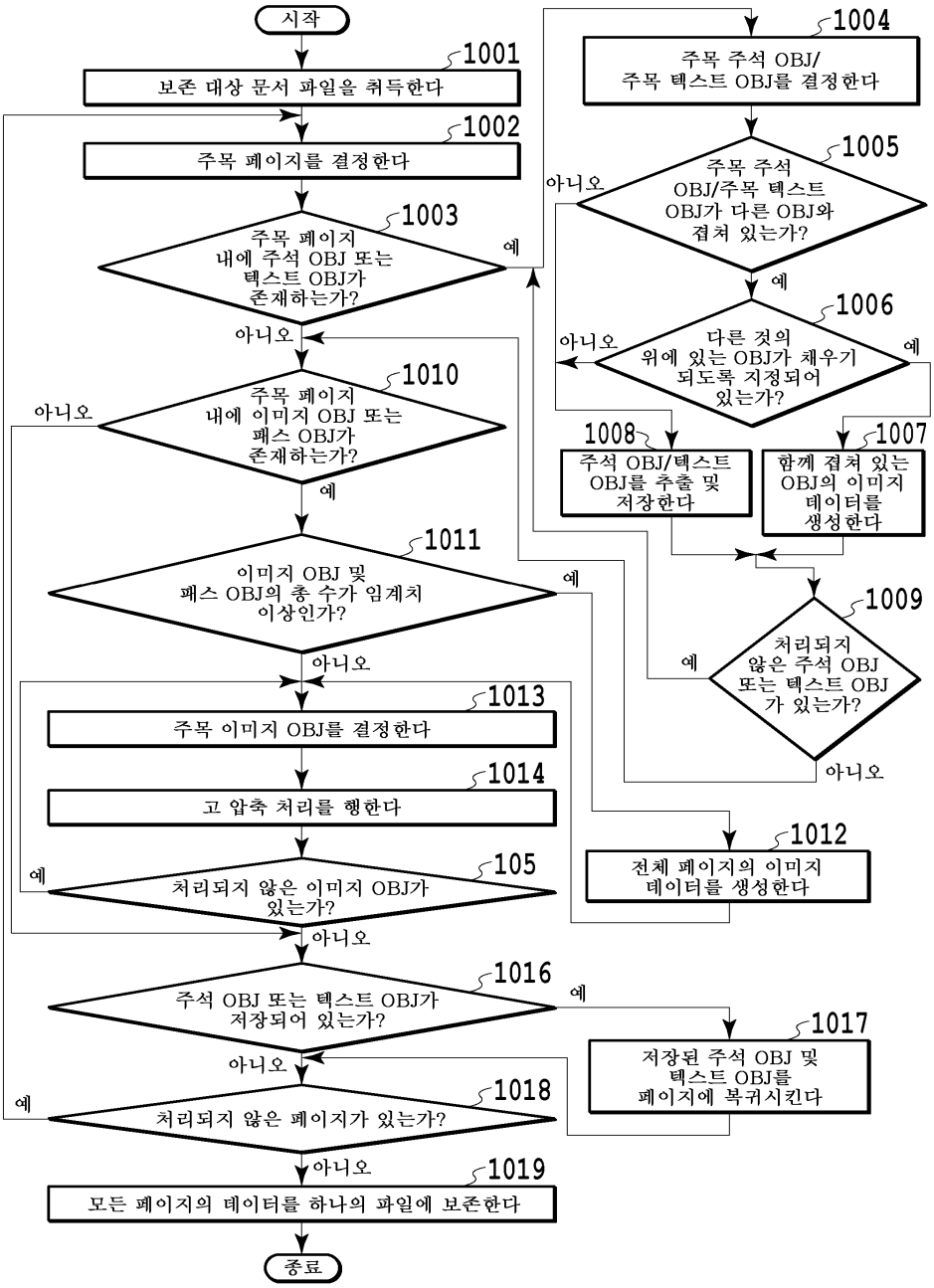
도면8


시작
801
보존 대상 문서 파일을 취득한다
802
주목 페이지를 결정한다
803
주목 페이지 내에 주석 OBJ 또는 텍스트 OBJ가 존재하는가?
예
804
주석 OBJ 및 텍스트 OBJ를 추출 및 저장한다
아니오
805
주목 페이지 내에 이미지 OBJ 또는 패스 OBJ가 존재하는가?
아니오
예
806
이미지 OBJ 및 패스 OBJ의 총 수가 임계치 이상인가?
예
807
주목 페이지 내에 공유 화상 OBJ가 존재하는가?
아니오
아니오
예
808
공유 이미지 OBJ를 제외한 전체 페이지의 이미지 데이터를 생성한다
809
전체 페이지의 이미지 데이터를 생성한다
810
주목 이미지 OBJ를 결정한다
811
고 압축 처리를 행한다
812
처리되지 않은 이미지 OBJ가 있는가?
예
아니오
813
주석 OBJ 또는 텍스트 OBJ가 저장되어 있는가?
예
814
저장된 주석 OBJ 및 텍스트 OBJ를 페이지로 복귀시킨다
아니오
815
처리되지 않은 페이지가 있는가?
예
아니오
816
모든 페이지의 데이터를 하나의 파일에 보존한다
종료

도면9



도면10


시작
1001
보존 대상 문서 파일을 취득한다
1002
주목 페이지를 결정한다
1003
주목 페이지 내에 주석 OBJ 또는 텍스트 OBJ가 존재하는가?
예
아니오
1004
주목 주석 OBJ/ 주목 텍스트 OBJ를 결정한다
1005
주목 주석 OBJ/주목 텍스트 OBJ가 다른 OBJ와 겹쳐 있는가?
아니오
예
1006
다른 것의 위에 있는 OBJ가 채우기 되도록 지정되어 있는가?
아니오
예
1008
주석 OBJ/텍스트 OBJ를 추출 및 저장한다
1007
함께 겹쳐 있는 OBJ의 이미지 데이터를 생성한다
1009
처리되지 않은 주석 OBJ 또는 텍스트 OBJ가 있는가?
예
아니오
1010
주목 페이지 내에 이미지 OBJ 또는 패스 OBJ가 존재하는가?
아니오
예
1011
이미지 OBJ 및 패스 OBJ의 총 수가 임계치 이상인가?
예
아니오
1012
전체 페이지의 이미지 데이터를 생성한다
1013
주목 이미지 OBJ를 결정한다
1014
고 압축 처리를 행한다
105
처리되지 않은 이미지 OBJ가 있는가?
예
아니오
1016
주석 OBJ 또는 텍스트 OBJ가 저장되어 있는가?
예
아니오
1017
저장된 주석 OBJ 및 텍스트 OBJ를 페이지에 복귀시킨다
1018
처리되지 않은 페이지가 있는가?
예
아니오
1019
모든 페이지의 데이터를 하나의 파일에 보존한다
종료