## **Automatic content extraction in document images**

# Nicola Carkaxhija Dept. of Information Engineering, University of Florence Via di S. Marta 3, 50139 – Florence, Italy

nicola.carkaxhija@stud.unifi.it

#### **Abstract**

In this paper a revisited version of an automatic document content extraction method [1] is proposed. The algorithm relies on an image registration procedure, that in a first stage enables the definition of a probabilistic template from a set of scanned images, then it matches incoming images to the built model to locate the positions where variable fields appear. The approach is validated on collections of forms filled in different ways to observe how results vary with them.

#### 1. Introduction

Despite the rise of the digital era, currently incoming data format of several entities still consists in paper. Migration of information stored in physical paper towards an electronic environment presents several advantages: storage, management, consultation, remote accessibility and security. Manual processing is a costly task and the presented method aims to make it automatic, eliminating the need of human intervention; this is particularly useful in historical collections where the template of the form is usually not available.

Information extraction is enabled by defining a template that predisposes documents to be read by an Optical Character Recognition (OCR) program, allowing to point out locations to be interpreted to extract the filled fields: indeed, the knowledge of the static part of a document allows, by contrast, to isolate the relevant content, that is represented by those portions that change over the instances. Thus, given a sufficiently large set of samples belonging to the same class, a fully automated process can generate a probabilistic template indicating the likelihood a pixel has to be either static or variable content.

Since the model is generated as average of many filled forms, with respect to it pixels in those parts that have been filled represent noise and thus may be interpreted as aleatory variables with a certain average removable by applying an opportune threshold.

In a summary, the method proposed in this paper consists of two different stages: in the former instances of the same document kind are provided to the system to generate a template, while in the latter the built model is used to process new incoming form images in order to extract the filled parts. Both the steps rests on an alignment algorithm illustrated hereinafter, that constitutes the main difference with respect to the original implementation to which the rest of the procedure abides.

### 2. Image registration

Given two instances belonging to the same class, image alignment is the task of finding the parameters that warp an image so that it minimizes the difference with the other one according a certain distance measure. In the case of a set with higher cardinality, it results in the representation of all the images into the same coordinate system.

To carry out this operation the homography estimation has been used instead of computing the Lukas and Kanade optical flow [2] as proposed in the original implementation of this content separation method, whereby several preprocessing steps are needed before alignment to make the algorithm attain more robustness, since otherwise it's sensible to local minima. Thus, the following operations became superfluous:

- connected component analysis to cut out irrelevant small regions too big ones that likely constitute margins of another document;
- the gradient module over the image to reduce the harmful contribution provided by too large foreground regions in the parameter estimation, potential cause of misalignments;
  - the consequent binarization with the Otsu algorithm [3];
- the typical multiscale approach used to face out the algorithm sensitivity to local minima and avoid to get stuck before obtaining the actual wrapping parameters.

Furthermore, the use of homographies as wrapping functions makes possible to cope with perspective distortions introduced when the scanner cover doesn't completely lower, entailing a non-uniform pressure on the document during the acquisition, rather than handling just images related at most by an affine transform: scaling, translation, skewing and rotation.

With regard to the error to minimize, a much more robust method as RANSAC has been adopted in lieu of the squared sum to cope with the presence of outliers deriving as a side effect of the removal of the many expedients above-mentioned.

When loading the images, they are still processed using the adaptive binarization algorithm proposed by Sauvola [4] to lessen the effects of an uneven illumination and reduce amount of points involved in the computations.

#### 3. Experimental results

The method has been tested on three document collections: the public NIST Structured Tax Forms Dataset (SPDB2) [5], whose fields has been filled through typewriters, a set of invoices with handwritten content and one of printed bills.

Many tries have been performed with sets of increasing size, but regardless the dataset involved, it could be claimed that a few dozen of samples are enough to produce a good template and the use of additional ones doesn't lead to perceptible improvements.

The best results have been obtained on the hand-filled invoices, foreseeable since such a content is characterized by a high variability.

The procedure yields good result also on NIST; however, not all the extracted fields are perfectly readable: specifically, this looks to be related to the use of a typewriter, that accurately centered at the begin of the blank space it's going to press on makes the first character camouflage as part of the template; because of the incline of the paper with respect to the machine this effect reduces as a long line is written.

In both the previous cases, content overlapping the template or too close to it is inevitably erased in the extraction phase. Instead, print-filled invoices characters are always located in fixed position and assimilated by the template during its generation, so that the extraction fails and the content results in a few small sparse spots. In case of thin printed content printing, such a problem could be partially overcome by selecting images randomly, in order to avoid influences by document acquisition order or nomenclature: e.g. in consecutive documents, a field assuming the same value may be interpreted as part of the template.

However, while dealing with different datasets, it emerged that the generation of an almost perfect model is possible just by a regularizing intervention, therefore the method is not completely automatic: when increasing the contrast by applying a sigmoid over the image is necessary to adjust the grey-level intensity on which it is centered; to make the procedure really unsupervised such a value must be made adaptive. For this purpose, an Otsu binarization has been performed, but it brought as a side-effect a slight degradation in the extraction result.

#### 4. Conclusions

By a qualitative analysis, the results achieved are satisfying, despite of some eroded characters that most of the times still are perceptible; in no way the method can suit to print-filled documents because unable to distinguish template and content.

The choice to don't use the Lukas and Kanade's algorithm in favour of a shorter and then simpler implementation worked, but it wasn't enough to reach a good result: there has been the need to carry out a further study to make the procedure effectively automatic.

## References

- [1] D. Aldavert, M. Rusinol and R. Toledo, "Automatic Static/Variable Content Separation in Adiministrative Document Images", in 2017 14th IAPR International Conference on Document Analysis and Recognition.
- [2] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings* of the 7<sup>th</sup> International Joint Conference on Artificial Intelligence Volume 2, 1981, pp. 674–679.
- [3] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, January 1979.
- [4] J. Sauvola and M. Pietikainen, "Adaptive document image binarization," "Pattern Recognition, vol. 33, no. 2, pp. 225–236.
- [5] D. Dimmick, M. Garris, and C. L. Wilson, "Structured forms database," National Institutte of Standards and Technology.

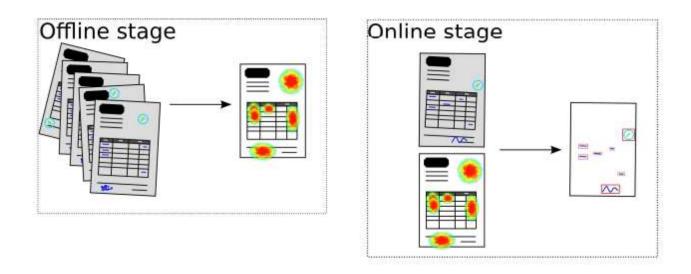


Figure 1: System overview, consisting in two phases: template generation and content extraction.

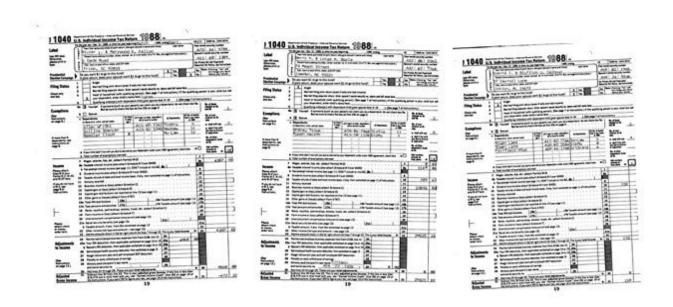


Figure 2: Samples of filled documents (with the same format) used to build the template.

<u>! 104(</u>	) $\tilde{\mathbf{u}}$	.S. Individual incom	rnal Revenue Se ne Tax Re	turn 19	38	XO.	•					
		or the year Jan -Dec 31, 1968, or	other tax year beg	inning	. 1988. •			. 19	OMB No. 154	15-0074		
Label	1	City, town or post office, state, and ZIP code  City. Town or post office, state, and ZIP code							ocial security num			
Use IRS label, Otherwise, please print or type.									Spouse's social security number  For Privacy Act and Paperwork Reduction Act Notice, see Instruction			
Presidential												
Election Campai	en 🕨	Do you want \$1 to go to the if joint return, does your s	is fund?	1 to go to this fu	nd?	Yes	/////	No No	Note: Checking "Y not change y reduce your	es" will our tax or		
Filing Status Check only one box.	1 2 3 4	Single Married filing joint Married filing separa Head of household your dependent, et	return (even if ite return. Enter d (with qualifyin nter child's nam	only one had incom spouse's social securi ng person). (See pa ne here	e) ty no. abov ge 7 of In:	structions.) If	the qualif	ying per	son is your child			
			such as your or	ent child (year spou	se died >	19 ) (S	e page 7	f Instru	ctions.)			
Exemptions (See		But be sure	to check the bo	x on line 33b on pag	e 2	ndent, do not	check box	5.	No. of boxes checked on Se and Sb			
Instructions on page 8.)	-	Dependents: (1) Name (first, initial, and last name)	(2) Check if under age 5	(3) If age 5 or older, depo social security numb		(4) Relationship	(5) No. of m Irved in your in 1981	home	No. of your children on Sc who:			
			_	<del></del>			ļ		e lived with you			
If more than 6 dependents, see Instructions on page 8.							İ		<ul> <li>didn't five with you due to divorce or separation</li> </ul>			
									No. of other dependents listed on Sc			
	•	If your child didn't live with you be Total number of exemptions					k here . I	• 🗆	Add numbers entered on	$\Box$		
	0.00	Wages, salaries, tips, etc. (at		(-2)		<del></del>		7	lines above	<u> </u>		
Income		Taxable interest income (also				· · · · · ·		B.		<del>  -                                    </del>		
Please attach		Tax-exempt interest income (s										
Copy B of your Forms W-2, W-2G,	9	Dividend income (also attach		•								
and W-2P here.	10	Taxable refunds of state and loc	ions	10								
If you do not have a W-2, see	12	Alimony received		12								
page 6 of	13	Capital gain or (loss) (attach 5		D				13				
Instructions.	14	Capital gain distributions not		13/see 03/9 11)				14		-		
	15	Other gains or (losses) (attack						15				
	160	1				amount (see )	111	166		-		
	17a	ī				amount (see		17b				
	18	Rents, royalties, partnerships	estates, trusts	etc. (attach Sched	lule F)	uniount (see )	rage 12)	18		_		
l	19	Farm income or (loss) (attach		19		ד						
	20	Unemployment compensation	(insurance) (s	ee page 13)		V 2 1 2		20				
Please attach check	21.	Social security benefits (see p	age 13)		210							
or money		Taxable amount, if any, from t						216				
order here.	22	Other income (list type and ar	nount—see pa	ge 13)				22				
		Add the amounts shown in the f				ur total incom	• . ▶	23				
Adimatmanta	24	Reimbursed employee business								C Central		
Adjustments		Your IRA deduction, from app					-		ĺ			
to Income		Spouse's IRA deduction, from a			256				ļ			
	26	Self-employed health insurance de			26							
···	27	Keogh retirement plan and sel			27							
(See Instructions	28	Penalty on early withdrawal of			28				1			
on page 13.)	29	Alimony paid (recipient's last r	name		Millitin							
	30	and social security no.		. ()	29				1			
Adjusted Gross Income	31	Add lines 24 through 29. Thes Subtract line 30 from line 23. \$18,576 and a child lived wit the Instructions. If you want life	This is your at	djusted gross incom			an of	30				
		The state of the s	o to the your	ies, see page 10 0	ine instru	ctions	. •	31				

				ua Sarvice	4000						
THAIL		individual (	income Ta	( HOTU	m 118100	(X) ending		. 1		OMB No. 1545-00	74
Fo	or the					1	Last name		our social	security number	-
	$\cdot$	Your first name an	A PURPLY (4) Notes carn	(ii, sizo g). i					A59	02 194	8_
bel [	Sierra I. & Bluffton U. Calhoun Present home address (number, street, and apt. no. or nural route). (If a P.O. Box, see page 6 of instruction							8.)		ocial security num	
RS label.	Ĭ Ì	Present home addr	A99	26 93	<u></u>						
herwise,	- 1	7 Carro		For Privace	Act and Paperwork						
ase print or	City, town or post office, state, and ZIP code									et Notice, see instr	Kliens
	: \	Cotton.	AL 14673			1 30	Yes		No Not	e: Checking "Yes"	will
	_			?		10	Yes		No	not change your reduce your ref	ind.
esidential ection Campaign	, II	oint return, do	es your spouse	want \$1	to go to this fund?.	-i A	165 1600		110		
	1	Single		(even if or	nly one had income)						
ling Status	2	X Married	filing joint return	Cotor or	ouse's social security no.	above am	full name h	ere		to come ability b	
	3	Married	filling separate retu	m. Eliter op	ouse's social security no. person). (See page 7	of Instru	ctions.) If ti	he qualify	ing perso	n is your child t	UL 170
heck only ne box.	4									>	
NE GOLA-		your de	pendent, enter co	decende	nt child (year spouse di	ed ▶ 19	). (See	page 7 o			
	5	Qualify	ing widow(er) with	Depende	rent) can claim you as a	depende	ent, do not o	theck box	6a. )	No. of boses	
	60	Yourself I	f someone (such	as your pai	on line 33b on page 2.				. }	No. of boxes checked on 6a and 6b	2
xemptions			But be sure to che								
See	b	Spouse .	<u> </u>		(3) If age 5 or older, dependen	r. T		(5) No. of m	oriks kome	No. of your children on 6c	
instructions	-	Dependents:		(2) Check	(3) If age 5 or coost, dependent social security number	(40)	Relationship	fixed in your		who:	
n page 8.)		(1) Name (first, initial,	, sed last name)	age 5		160	1 24	5		e lived with you	_4
		Rowan La			A11: 02: 778	16 +	Mo	0		- distribution with	
		Rowan La	ke		A33: 02: 558	4 5 4	Crand	6		you due to divorce or separation	
		Fairfie	1d Hood		A39: 64: 176	יוטוש	di dila	+			
If more than 6		F GITTIO								No. of other dependents listed on Sc	
dependents, see					i i .			+		on Sc	_
Instructions on page 8.				1	1 1				▶Ū	Add numbers	_
pege o.				atalanad ar	your dependent under a pr	e-1985 ag	reement, che	ck here .		entered on lines above	1.6
	•	If your child didn't	t live with you but is	COMMITTED 45	ou dependent				-	MING 61010	┰
									12		+-
	7	Wages salaries	s, tips, etc. (attaci	h Form(s)	W-2)				. 80		+
								1	2000		1
income	-								. 9		+
Please attach				had de A	f over \$400) · · ·				10	158	31
		Tax-exempt in	ma (elen attach Sc								
Copy B of your	,	Dividend incom	me (also attach So	income tax	es, if any, from workshe	et on page	11 of instri	uctions .	11		1
Copy 8 of your Forms W-2, W-2G, and W-2P here.	10	Dividend incom Taxable refunds	me <i>(also attach Sc</i> is of state and local	income ta	rover 3400) ies, if any, from workshe	et on page	11 of instru	uctions .			-
and W-2P here.	•	Dividend incor Taxable refund	me (also attach So is of state and local	income ta:	es, ir arry, from works	et on page	11 or instru	uctions .	11		-
and W-2P here. If you do not have	10	Dividend incom Taxable refunds Alimony receiv	me (also attach Sc is of state and local wed	h Schedu	le C).	et on pag	11 or instru	uctions .	11 12 13		-
and W-2P here. If you do not have	10	Dividend incom Taxable refunds Alimony receiv Business incom	me (also attach Sc is of state and local wed	:h Schedu	le C).	et on pag	11 or instru	uctions .	11 12 13 14		
and W-2P here. If you do not have	10 11 12	Dividend incom Taxable refunds Alimony receiv Business inco Capital gain of	me (also attach Sc is of state and local wed	th Schedu hedule D)	le C).				11 12 13 14		
and W-2P here. If you do not have	10 11 12	Dividend incom Taxable refunds Alimony receiv Business inco Capital gain of Capital gain of	me (also attach Sc is of state and local yed	th Schedu hedule D) ported on Form 479	le C)	b Taxabit	amount (s	ee page 1	11 12 13 14 15 1) 169		
and W-2P here.  If you do not have	10 11 12 13 14	Dividend incom Taxable refunds Alimony receiv Business inco Capital gain of Capital gain of Other gains of	me (also attach Sc is of state and local wed	th Schedu hedule D) ported on Form 479	le C)	b Taxabit	amount (s	ee page 1	11 12 13 14 15 1) 16b 2) 17b		
and W-2P here. If you do not have	10 11 12 13 14 15	Dividend incom Taxable refunds Alimony receiv Business inco Capital gain of Capital gain of Other gains of Total IRA dist	me (also attach Sc is of state and local wed me or (loss) (attach & (loss) (attach Sc listributions not re or (losses) (attach Iributions 12	th Schedu hedule D) ported on Form 479	line 13 (see page 11)	b Taxabi		ee page 1	11 12 13 14 15 1) 169		
and W-2P here. If you do not have	10 11 12 13 14 15 16	Dividend incommendation of the control of the contr	me (also attach Sc is of state and local ved	h Schedu hedule D) ported on Form 479	le C). line 13 (see page 11) 7) 16 17 18 18 19 19 19 19 19 19 19 19 19 19 19 19 19	b Taxabi	amount (s	ee page 1	11 12 13 14 15 1) 16b 2) 17b		
and W-2P here. If you do not have	10 11 12 13 14 15 16 11	Dividend incommended in a control of the control of	me (also attach SC s of state and local ved me or (loss) (attach sc (loss) (attach Sc sistributions not re or (losses) (attach cributions 12 s and annuities lies, partnerships.	th Schedule D) ported on Form 479 Sa   estates, tr	line 13 (see page 11) 7) 16 17 17 18 18 18 18 18 18 18 18 18 18 18 18 18	b Taxabi	amount (s	ee page 1	11 12 13 14 15 1) 16b 2) 17b 18		
and W-2P here. If you do not have	10 11 12 13 14 15 16 17	Dividend incommended in a control of the control of	me (also attach SC s of state and local ved me or (loss) (attach sc (loss) (attach Sc sistributions not re or (losses) (attach cributions 12 s and annuities lies, partnerships.	th Schedule D) ported on Form 479 Sa   estates, tr	line 13 (see page 11) 7) 16 17 17 18 18 18 18 18 18 18 18 18 18 18 18 18	b Taxabi	amount (s	ee page 1	11 12 13 14 15 1) 169 2) 178 18 19		
and W-2P here. If you do not have	10 11 12 13 14 11 11 11	Dividend incommended in a commendation of the	me (also attach SCs is of state and local vetoes of closs) (attach SC istributions not re ricloses) (attach SC istributions not re ricloses) (attach tributions	th Schedule D) sported on Form 479 [Em ] estates, tr Schedule II (insurance	line 13 (see page 11) 7) 16 17 17 18 18 18 18 18 18 18 18 18 18 18 18 18	b Taxabi	amount (s	ee page 1	11 12 13 14 15 10 16b 17b 18 19 20		
and W-2P here. If you do not have a W-2, see page 6 of instructions.	10 11 12 13 14 11 11 11	Dividend incom Taxable refunds Alimony receiv Business inco Capital gain of Capital gain of Capital gain of Total pensions Rents, royalti Farm income Unemployme	me (also attach SCs so I state and local weed me or (loss) (attach Sc listributions not re rr (losses) (attach a and annuties is, partnerships, e or (loss) (attach ent compensation in honefits (see B.	th Schedule D) ported on Form 479 (6a ) restates, to Schedule I. (insurance age 13)	tine 13 (see page 11)  11  12  15  17  17  18  19  19  19  19  19  19  19  19  19	b Taxabi	amount (s	ee page 1	11 12 13 14 15 10 16b 20 17b 18 19 20		
end W-2P here. If you do not have a W-2, see page 6 of instructions.  Please attach check	10 11 12 13 14 11 11 11 12 2	Dividend incom Taxable refunds Alimony receiv Business inco Capital gain of Capital gain of Capital gain of Other gains o Sa Total pensions Rents, royalti Farm income O Unemployme La Social securi Taxable amo	me (also attach SCs of size and local sed of size and local me or (loss) (attach SC size its ributions not re it (losses) (attach Cributions in the tributions in the sand annuities. List, and annuities, partnerships, eer (loss) (attach ent compensation ity benefits (see punt, if any, from it	income tai ch Schedu hedule D) ported on Form 479 (Sa     17a   estates, ti Schedule I (insurance age 13) the worksh	(e C).  line 13 (see page 11)  7)  16  17  18  18  19  19  19  19  19  19  19  19	b Taxabii b Taxabii dule E)	e amount (s	ee page 1 ee page 1	11 12 13 14 15 10 16b 17b 19 20		
and W-2P here. If you do not have a W-2, see page 6 of instructions.	10 11 12 13 14 15 16 17 17 18 18 18 18 18 18 18 18 18 18 18 18 18	Dividend incon Taxable refund Alimony recels Business inco Capital gain or Capital gain or Control and Cother gains or Rents, royalti Farm income Unemployme Social Securit Taxable amo Unter income Unterployme Other income	me (also attach SCs of size and local wed (loss) (attach SC (loss) (attach SC (loss)) (attach SC (losses) (attach Intibutions not re in (losses) (attach Intibutions List sand annualize sand annualize see and annualize heart compensation ity benefits (see punt, if any, from it et (list type and are to the see (list type and are type are to the see (list type and are type	income tai th Schedule D) iported on Form 479 (&     17a   estates, tr Schedule I (insurance age 13) the workst nount—Sear right sol	tine 13 (see page 11)  7)  16  17  18  19  19  19  19  19  19  19  19  19	b Taxabii b Taxabii dule E)	e amount (s	ee page 1 ee page 1	11 12 13 14 15 10 16b 20 17b 18 19 20		
end W-2P here.  If you do not have a W-2, see page 6 of instructions.  Please attach check or money	10 11 12 13 14 15 16 17 17 18 18 18 18 18 18 18 18 18 18 18 18 18	Dividend incon Taxable refund Alimony recels Business inco Capital gain or Capital gain or Control and Cother gains or Rents, royalti Farm income Unemployme Social Securit Taxable amo Unter income Unterployme Other income	me (also attach SCs of size and local wed (loss) (attach SC (loss) (attach SC (loss)) (attach SC (losses) (attach Intibutions not re in (losses) (attach Intibutions List sand annualize sand annualize see and annualize heart compensation ity benefits (see punt, if any, from it et (list type and are to the see (list type and are type are to the see (list type and are type	income tai th Schedule D) iported on Form 479 (&     17a   estates, tr Schedule I (insurance age 13) the workst nount—Sear right sol	tine 13 (see page 11)  7)  16  17  18  19  19  19  19  19  19  19  19  19	b Taxabii b Taxabii dule E)	e amount (s	ee page 1 ee page 1	11 12 13 14 15 10 16b 17b 19 20		
end W-2P here.  If you do not have a W-2, see page 6 of instructions.  Please attach check or money	10 11 12 13 14 11 10 11 11 12 2	Dividend incor Taxable refund Alimony receiv Business incor Capital gain of Capital gain of Other gains or Total IRA dist Ta Total pensions Rents, royahi Farm income Unemployme Taxable amo Unterployme Taxable amo Add the amor	me (also attach SCs is of state and local wed or (loss) (attach sc (loss) (attach sc (loss) (attach sc (loss) (attach sc (loss)) (attach sc (l	income tai th Schedule D) ported on Form 479 See   17a   estates, to Schedule I (insurance age 13) the workst four right con	tine 13 (see page 11) 7) 16 17 17 18 19 19 10 19 10 10 10 10 10 10 10 10 10 10 10 10 10	b Taxabil b Taxabil dule E)	e amount (s	ee page 1 ee page 1	11 12 13 14 15 10 16b 17b 19 20		
and W-2P here. If you do not have a you are been a for instructions.  Please attach check or money order here.	10 11 12 13 14 11 10 11 11 12 2	Dividend incon Taxable refundi Alimony recels Business incor Capital gain or Total RA data Total pensions Rents, royalti Farm income Unemployme 1a Social securi Taxable amo O Uner incom Add the amo Add the amo Reimbursed	me (also attach SCs is of state and local wed (loss) (attach sc (loss) (attach sc (loss) (attach sc (loss)) (attach sc (los	income tai th Schedule D) ported on Form 479 Sa   (7a   estates, ti Schedule I (insurance age 13) the workst nount—si ar right col ; expense wi	line 13 (see page 11) 7) 16 16 17 18 19 19 19 19 19 19 19 19 19 19 19 19 19	b Taxabin b Taxabin dule E) 21a 22. This in 24	e amount (s	ee page 1 ee page 1	11 12 13 14 15 10 16b 17b 19 20		
and W-2P here. If you do not have a W-2, see page 60 instructions.  Please stach check or covery or dee here.  Adjustments	10 11 12 13 14 11 10 11 11 12 2	Dividend incom Taxable refundi Alimony recels Business inco Capital gain of Capital gain of Capital gain of Other gains or a Total IRA dist Total pensions Rents, royalti Farm incom O Unemploym 1a Soxial securi Taxable arm O Under incom Add the amo 2 Other incom Add the amo 2 Heimbursed Reimbursed Reimbursed	is of state and local seed	income tai.  ch Schedule D) ported on Form 479 [Sa ] [7a ] estates, ti Schedule I (insurance age 13) the workst nount—sear right col expenses blicable we conticable to co	tine 13 (see page 11)  1 16  1 15  1 16  1 16  1 16  1 17  1 16  1	b Taxabin b Taxabin dule E) 22. This in 24 15 25s 15 25b	e amount (s	ee page 1 ee page 1	11 12 13 14 15 10 16b 17b 19 20		
and W-2P here. If you do not have a yea can be yease for instructions.  Please attach check or money order here.	100 111 122 133 144 151 161 171 171 172 222	Divided incor Taxabir erfulor Alimony recels Business inco Capital gain of Capital gain of Capital gain of Capital gain of Capital gain of Capital gain of Rents, royall Farm incommo Unemployme 1a Social securi b Taxable amo Unemployme 1a Social securi b Taxable amo 1a Add the amo 3. Add the amo 4. Reimbursed 5a Your IRA of Spouse's IRI	me (also attach Sc is of state and local wed . (loss) (attach Sc listributions not re in (losses) (attach sc listributions not re in (losses) (attach- ies, partnerships, e or (loss) (attach- ent compensation ity benefits (see p sunt, if any, from the (last type and ar unts shown in the I employee business	income tai  h Schedule D) ported on Form 479    Sa     17a     cinsurant age 13)   the worksh nount—star right col   expasses   cinsurant    te C).  line 13 (see page 11)  16  17)  16  17)  16  17)  16  17)  16  17)  18  18  19  19  19  19  19  19  19  19	b Taxabin b Taxabin dule E) 22. This in 24 15 25s 15 25b	e amount (s	ee page 1 ee page 1	11 12 13 14 15 10 16b 17b 19 20			
and W-2P here. If you do not have a W-2, see page 60 instructions.  Please stach check or covery or dee here.  Adjustments	100 111 122 133 144 111 111 112 22 23	Dividend incort Asimony receiv Business inco Capital gain of Capital gain Capital gain Capital gain Capital Ca	so state and local mere (also attach Sc so state and local mere (loss) (attach mere (l	th Schedule D) ported on Form 479 (6m   1/7a   estates, til Schedule I) (insurance age 13) the workshount—siar right col: expenses alicable we applicable with the firemploy of	tine 13 (see page 11) 7) 16 17 19 18 19 19 19 19 19 19 19 19 19 19 19 19 19	b Taxabid b Taxabid dule E) 21a 22. This id 3. 24 15 25a 26 27	e amount (s	ee page 1 ee page 1	11 12 13 14 15 10 16b 17b 19 20		
end W-2P here. H you do not have a W-2, see page 6 of instructions.  Please stach check or coory order here.  Adjustments	100 111 122 133 144 111 111 112 22	Dividend incort Asimony receiv Business inco Capital gain of Capital gain Capital gain Capital gain Capital Ca	so state and local mere (also attach Sc so state and local mere (loss) (attach mere (l	th Schedule D) ported on Form 479 (6m   1/7a   estates, til Schedule I) (insurance age 13) the workshount—siar right col: expenses alicable we applicable with the firemploy of	tine 13 (see page 11) 7) 16 17 19 18 19 19 19 19 19 19 19 19 19 19 19 19 19	b Taxabin b Taxabin dule E) 22. This in 24 15 25s 15 25b	e amount (s	ee page 1 ee page 1	11 12 13 14 15 10 16b 17b 19 20		
was W-2P here. If you do not have a W-2, see have sage 6 of instructions.  Please stach check of the was a w	100 111 122 133 144 151 161 171 172 222 233	Dividend incon Trasple refunda Alimony receiv Business incor Capital gain of Capital gain of Capital gain of Capital gain of Chite gains o gain so is Total IRA dist. Tar Total pensions a Total IRA dist. Tar Total pensions of Total IRA dist. Tar Total pensions of Total IRA dist. Tar Total pensions Total IRA dist. Tar Total pensions of Pensions Capital gain of Capital gain of Capital gain of Capital gain Total IRA dist. Tar. A Reimbursded A Reimbursded Total IRA dist. Tar. Capital Total	one (also attach Sc on state in Sc on St on State in Sc on State i	th Schedule D) ported on Form 479 (6m   1/7a   estates, to Schedule I) (insurance age 13) the workst mount—site right color expenses olicable we applicable with the feduction, fireff-employ of savings	tine 13 (see page 11) 7) 16 17 19 18 19 19 19 19 19 19 19 19 19 19 19 19 19	b Taxabin b Taxabin dule E)  21a  22 This in 24 15 25b 25b 26 27 28	e amount (s	ee page 1 ee page 1	11 12 13 14 15 10 16b 17b 19 20		58
and W-2P here. If you do not have a W-2, see here begans of instructions.  Please stach check or money order here.  Adjustments to Income	100 111 122 133 144 151 161 171 172 222 233	Dividend incont Trauble refund in Alimony received a fundamental part of the Alimony received a fundamental part of the Alimony received a fundamental part of the Alimony and	on sistem hose of state hose of state hose of state hose of state and coal state hose of stat	th Schedule D) ported on Form 479 file	tine 13 (see page 11)  16  16  16  16  16  16  16  16  16	b Taxabid b Taxabid dule E) 21a 22. This id 3. 24 15 25a 26 27	e amount (s	ee page 1 ee page 1	11 12 13 14 14 15 15 16b 22) 12b 20 211 22 21 22	15	
end W-2P here. If you do not have aw 2, see have age 6 of instructions.  Please stach check other article have goden hare.  Adjustments to income	100 111 122 133 144 151 161 171 172 222 233	Dividend incont Trashle refunda Alimony tecels Business incording to Capital gain of Capital g	on state in Science and a state in Science and a state in Science and consideration and a state and science and a state and science and a state and science and a state and a	th Schedu hedule D) ported on 79 [6m   17a   17a	tine 13 (see page 11)  ine 13 (see page 11)  ine 13 (see page 11)  ine 14 (stack Sche 7)  seet on page 13  seet on page 13  see page 13)  unto for iner 7 through  inchaheet on page 14  on morksheet on page 14  of Sche 7)  seed SEP deduction.	b Taxabli b Taxabli dule E) 22. This is 22. This is 15. 25. 26. 27. 28.	amount (se amount (se amount (se	ee page 1	11 12 13 14 15 15 19 20 21 22 22 3 4 19 20 21 22 22 3 4 4 15 15 15 16 16 16 17 18 18 18 18 18 18 18 18 18 18	115	ø
and W-2P here. If you do not have a W-2, see here begans of instructions.  Please stach check or money order here.  Adjustments to Income	100 111 122 133 144 151 161 171 172 222 233	Dividend incort Tasable refunding Alimony receive Dusiness incore Capital gain di Capital Capi	os o states não so os states não so os states não so os states não so os states não como os os states não sou o states não sou se os	th Schedu hedule D) ported on 79 (Sa   1/2 a	tine 13 (see page 11) 7) 16 17 19 18 19 19 19 19 19 19 19 19 19 19 19 19 19	b Taxabih dule E)  21e  22. This is 24  15 25e  27 28  29 ncome A	amount (se	ee page 1 ee page 1 noome	11 12 13 14 15 15 19 20 21 22 22 3 4 19 20 21 22 22 3 4 4 15 15 15 16 16 16 17 18 18 18 18 18 18 18 18 18 18	115	58

Figure 4: Test image and result of the extraction phase.

Sierra I. & Bluffto	on U. I	Calh	OLU			A59	02	1948
57 Carroll Lane						A99	26	9350
Chitton, AL 14673								
				X				
x								
x								
,								2
								-
Rowan Lake	A11	02	7786	So-Law	. 5			4
Rowan Lake Fairfield HPod	A33	02	5504	StpMo GrtGrand				
		-						
				-				
								6
							15	58
								Mari
							15	8
							158	2