

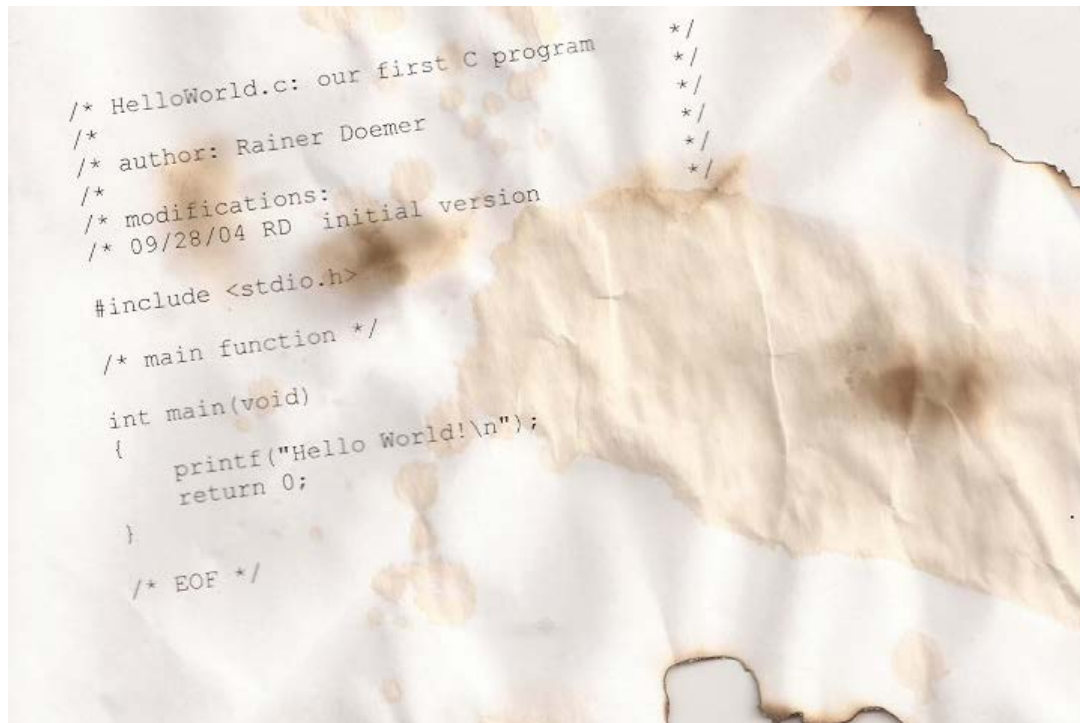


# **OCR** **(OPTICAL CHARACTER RECOGNITION)**

**Che-Wei Chang**

# OCR (OPTICAL CHARACTER RECOGNITION)

- This is the goal...



# OCR (OPTICAL CHARACTER RECOGNITION)

- Intuitive idea

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z						
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z						
1	2	3	4	5	6	7	8	9	0																						
~	!	@	#	\$	%	^	&	*	(	)	_	+	`	-	=	{	}		[	]	\	:	"	<	>	?	;	'	,	.	/

- All characters take the same space in Courier font.
- No overlapping between characters in Courier font.



# OCR (OPTICAL CHARACTER RECOGNITION)

- Intuitive idea

```
/* HelloWorld.c: our first C program */
/*
/* author: Rainer Doemer */
/*
/* modifications: */
/* 09/28/04 RD initial version */

#include <stdio.h>

/* main function */

int main(void)
{
    printf("Hello World!\n");
    return 0;
}

/* EOF */
```



# OCR (OPTICAL CHARACTER RECOGNITION)

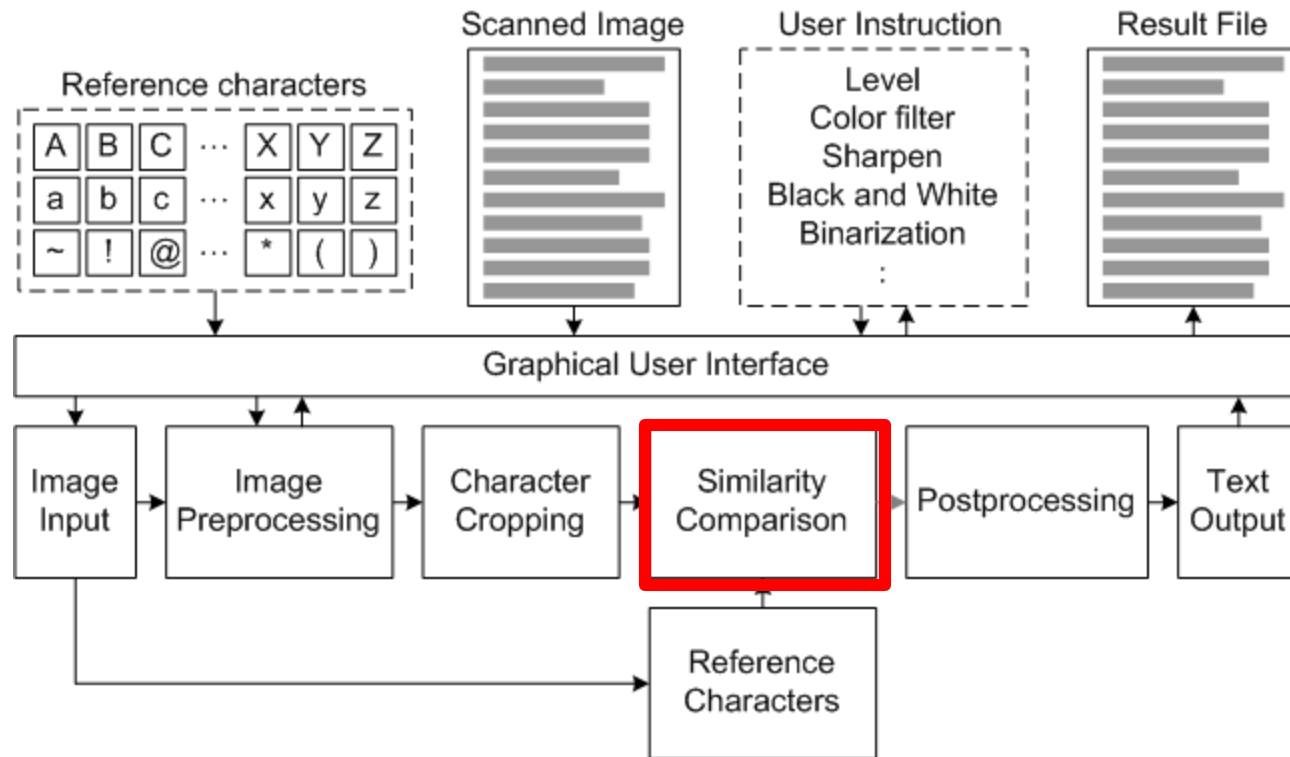
- Intuitive idea

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z						
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z						
1	2	3	4	5	6	7	8	9	0																						
~	!	@	#	\$	%	^	&	*	(	)	_	+	`	-	=	{	}		[	]	\	:	"	<	>	?	;	'	,	.	/

- Cropping the characters
- Matching the characters with the reference characters in the library.
- Matching the word with the key words of C language in the dictionary.

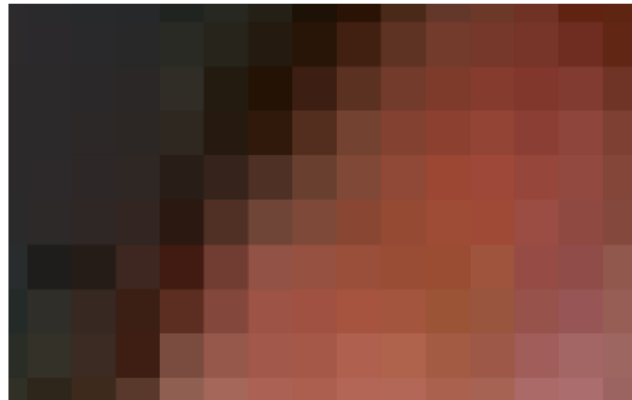
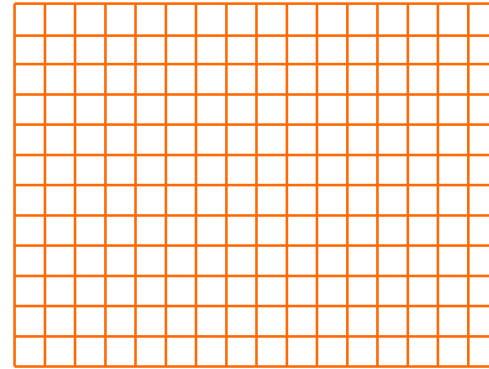


# OCR ARCHITECTURE



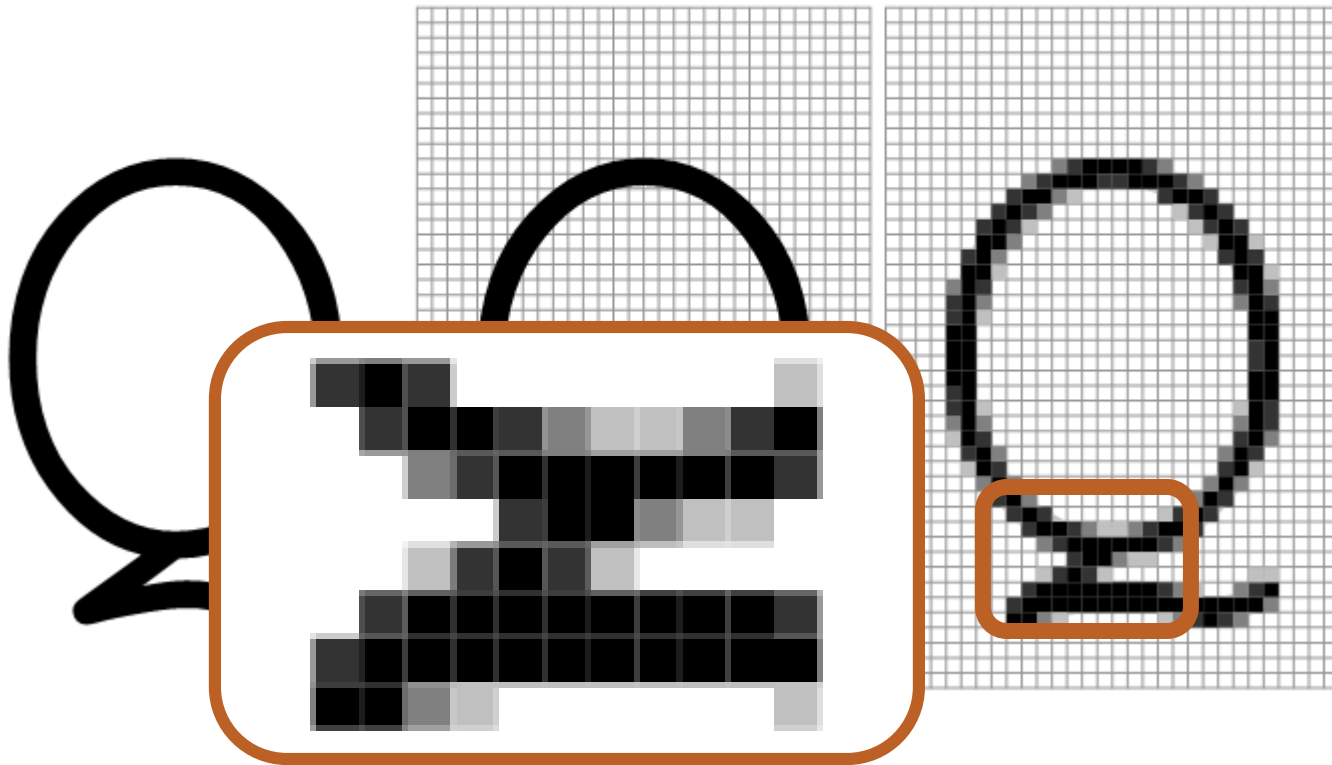
# PICTURE IN THE PROGRAM

- How to represent a picture in computer?
  - A picture is composed of pixels
  - One color for each pixel
  - Example:  $16 \times 12 = 192$  pixels



# CHARACTER MATCHING

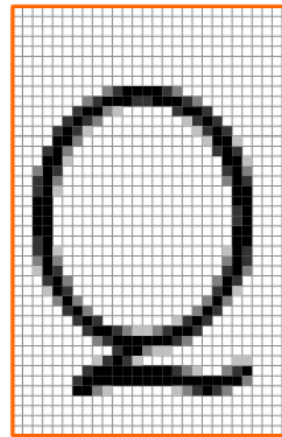
- Character “Q” in pixels (resolution 30x46)





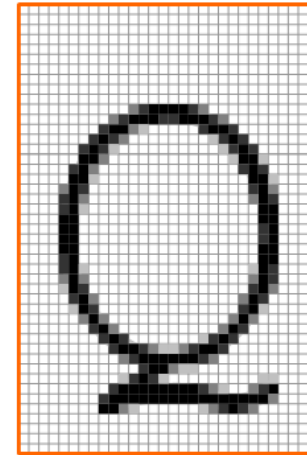
# ISSUES

- Size Mismatching



28

44



30

46

- Offset

