

# Análise Forense de Documentos Digitais

*Prof. Dr. Anderson Rocha*

[anderson.rocha@ic.unicamp.br](mailto:anderson.rocha@ic.unicamp.br)

<http://www.ic.unicamp.br/~rocha>

---

Reasoning for Complex Data (RECOD) Lab.  
Institute of Computing, Unicamp

Av. Albert Einstein, 1251 – Cidade Universitária  
CEP 13083-970 • Campinas/SP – Brasil

---

# **Identificação de** ***Scanners* e** **Impressoras**

# **Organização**

# Organização

## ▶ Introdução

## ▶ Impressoras (Laser e *Inkjet*)

- Funcionamento
- Fraudes
- Técnicas

## ▶ *Scanners*

- Funcionamento
- Técnicas
- Resultados

## ▶ Conclusão

## ▶ Referências

# Introdução

# Introdução

## ► Motivação

- Foram gastos **US\$633M/ano** em combate a pirataria nos anos de 1995-2001;
- Validar cópias legais;
- Falsificação de documentos (“*scan-print*”);

# Introdução

- ▶ Técnicas para identificação de impressoras e *scanners*  
[Chiang et al. 2009]
- ▶ Análise de ruído considerando as propriedades de captura unidimensionais e periódicas do *scanner*  
[Khanna et al. 2009]

# **Impressoras**

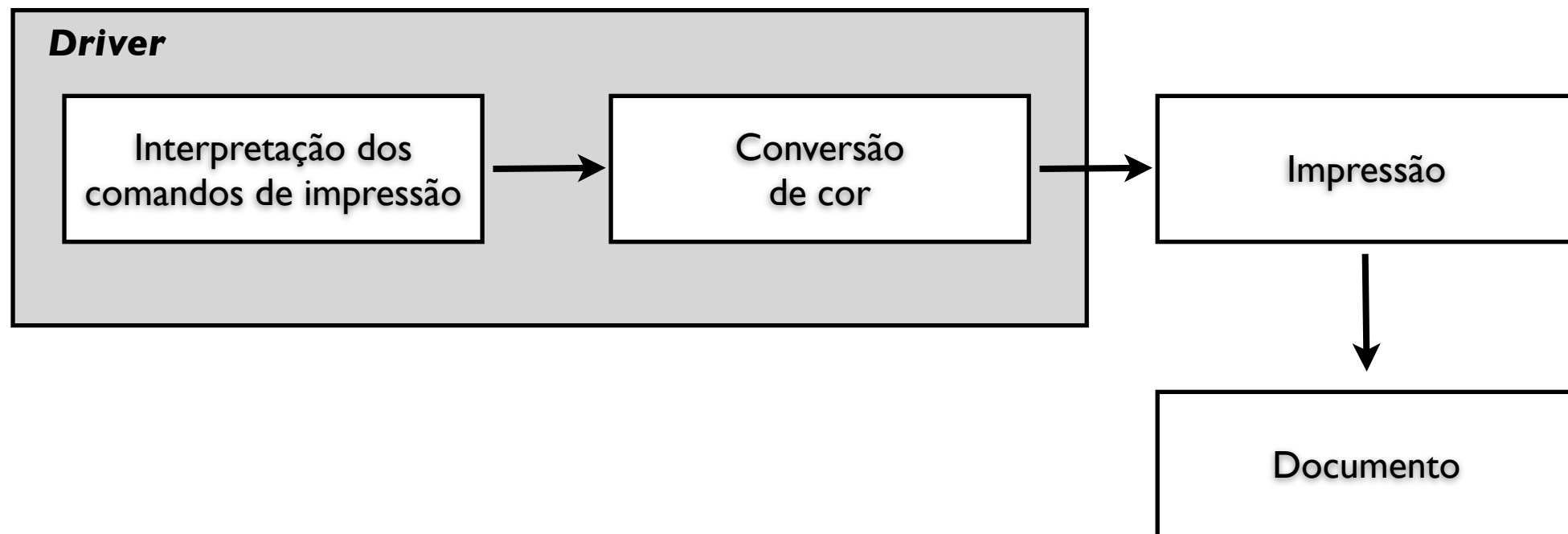


# Impressoras

- ▶ Características intrínsecas
  - Defeitos de fabricação (invisível ao olho humano)
- ▶ Características extrínsecas
  - Imprimir marca d'água

# Impressoras

## ► Pipeline de uma impressora



© M. Nagliati

# Impressoras

- ▶ Podemos inserir informação no documento, afim de usá-las como marca d'água;
- ▶ A primeira técnica é colocar informações no processo de interpretação dos comandos e conversão de cores;
  - Porém esse método é pouco utilizado, pois ele ocorre dentro do *driver*, e é fácil de desfazer.

# Impressoras

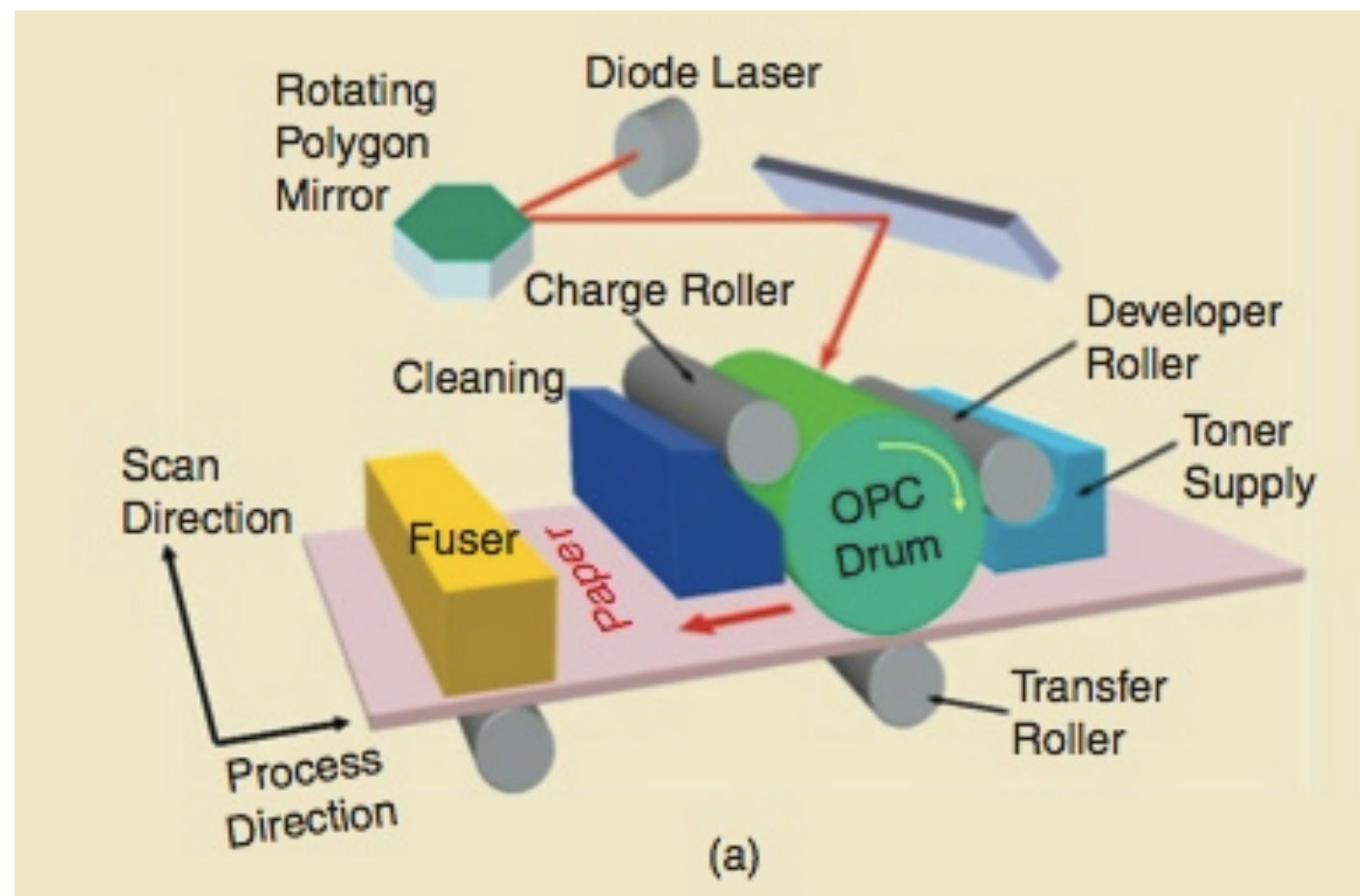
- ▶ Outra técnica, é inserir essas marcas d'água dentro do *hardware* da impressora;
  - Para fazer algum tipo de alteração, é preciso mexer no *firmware* da impressora.

# Impressora *Laser*

- ▶ *Electrophotographic printer*
- ▶ Rolo magnetizado com tinta (*tonner*)
- ▶ Diodo retira a carga em certos pontos (forma a imagem a ser impressa)
- ▶ Folha passa pelo rolo e absorve a tinta

# Impressora *Laser*

## ► Funcionamento



© P. Chiang

# Impressora *Laser*

- ▶ Como obter a “assinatura” de uma impressora *Laser*?
- ▶ Imperfeições nas engrenagens do sistema da impressora, geram um efeito periódico que pode ser explorado como uma assinatura;
- ▶ Esse efeito cria linhas (claras ou escuras) ao longo da impressão, e pode ser atribuído à uma única impressora.

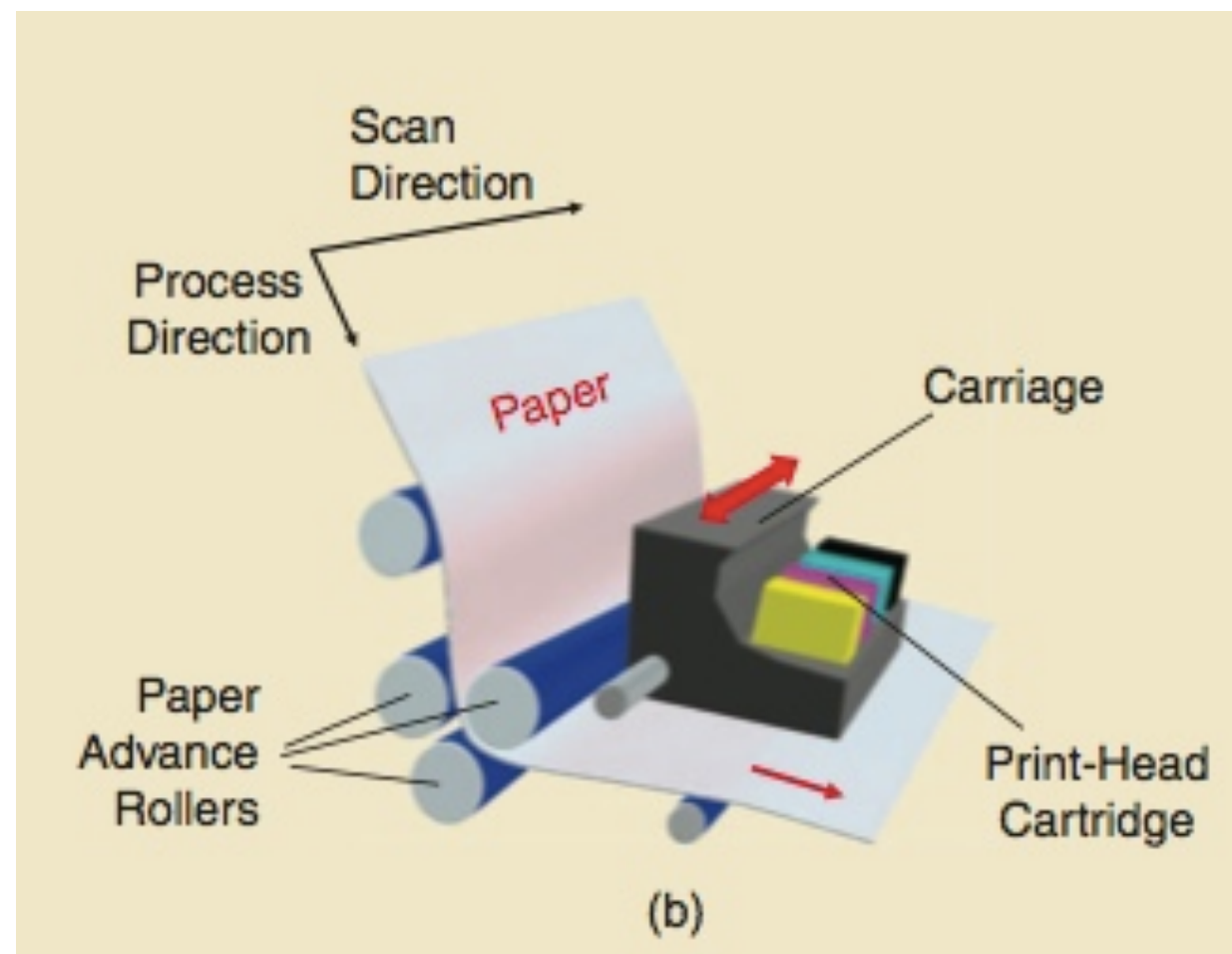
# Impressora *Inkjet*

- ▶ Três componentes principais
  - Cabeça de impressão
  - Cartucho
  - Mecanismo de avanço do papel



# Impressora *Inkjet*

## ► Funcionamento



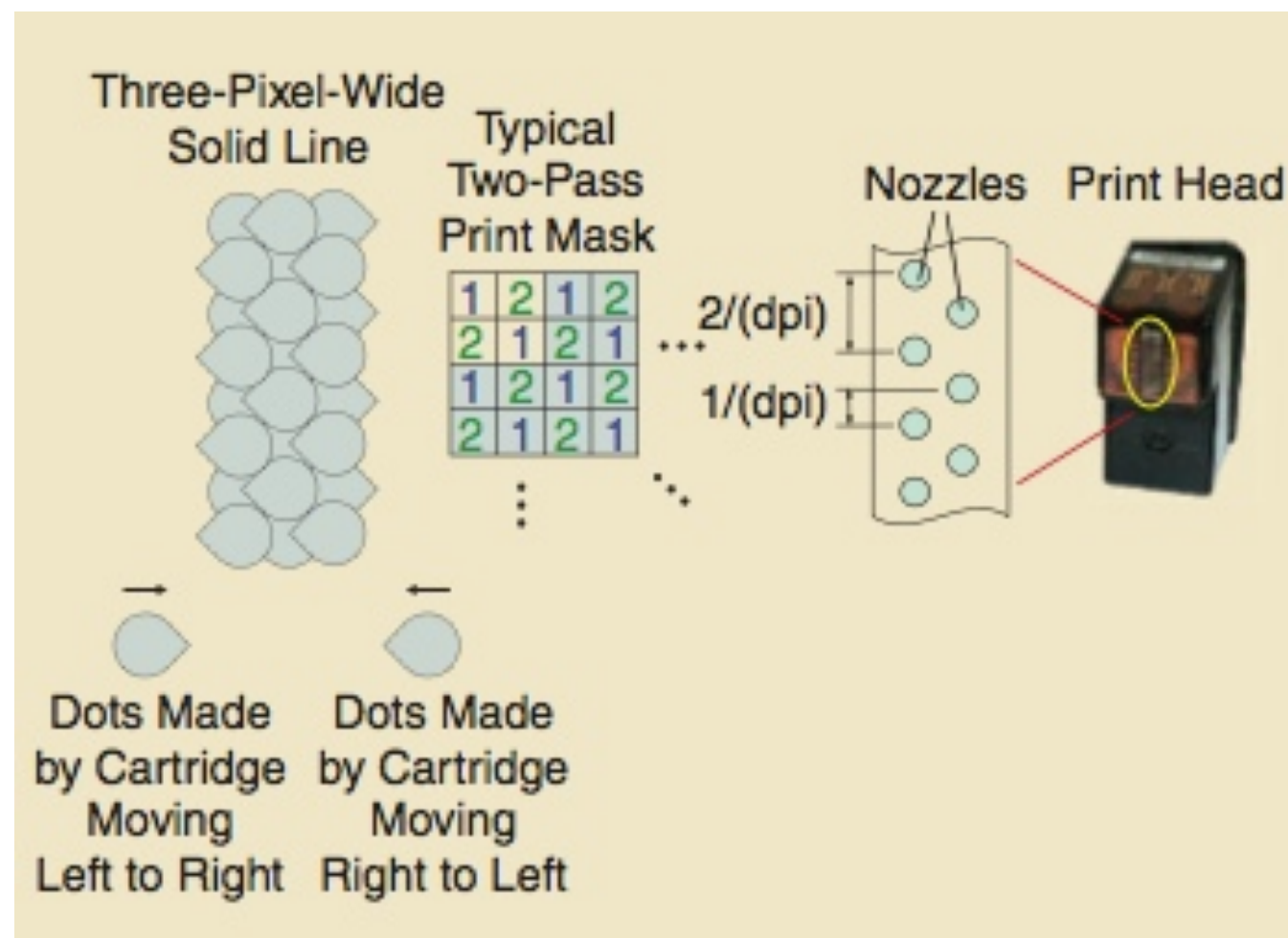
© P. Chiang

# Impressora *Inkjet*

- ▶ Funcionamento parecido com o do *scanner*
  - A cabeça move de um lado para o outro e o mecanismo de avanço do papel move o papel na direção perpendicular;
- ▶ Unidirecional X Bidirecional
  - Nas impressoras unidirecionais, a cabeça de impressão, libera tinta em um único sentido, já a bidirecional, em ambos;

# Impressora *Inkjet*

## ► Funcionamento



© P. Chiang

# Impressora *Inkjet*

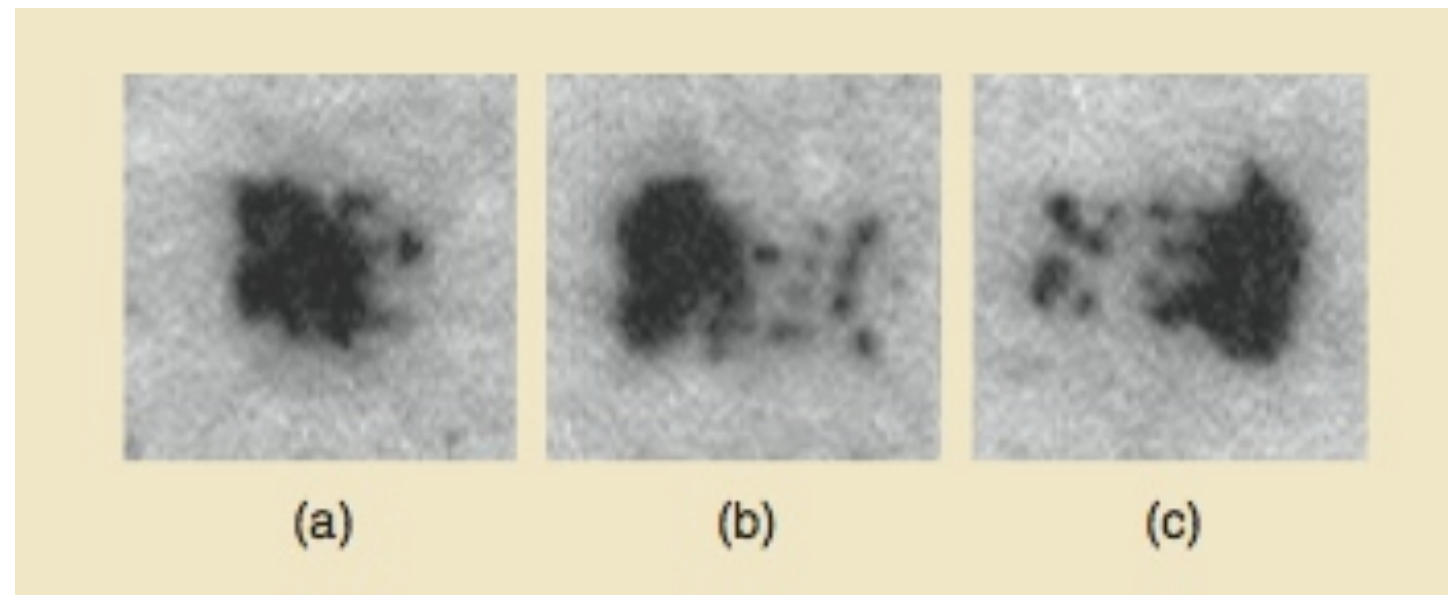
- ▶ A cabeça de impressão, possui pequenos furos, por onde sai a gota de tinta;
- ▶ Esse furos são localizados em duas linhas;
- ▶ Nas impressão bidirecional, existe sobreposição das gotas de tinta, pois o papel é movido de  $1/N$ , onde  $N$  é a largura da cabeça de impressão.

# Impressora *Inkjet*

- ▶ Como obter a “assinatura” de uma impressora *Inkjet*?
- ▶ Imperfeições no mecanismo de engrenagens;
- ▶ Química da tinta;
- ▶ Erro no avanço do papel;
- ▶ Falhas na cabeça de impressão;
- ▶ Imperfeições devido a velocidade de impressão.

# Impressora *Inkjet*

► Imperfeições devido a velocidade de impressão.



© P. Chiang

- a) 15 inch/sec, left-to-right
- b) 45 inch/sec, left-to-right
- c) 45 inch/sec, right-to-left

# Técnicas de Identificação

- ▶ Vamos apresentar duas técnicas para atribuir um documento à uma impressora específica;
  - Inserção de pontos amarelos no documento;
  - *Shift* de linhas e caracteres;

# Técnicas de Identificação

- ▶ Nas impressoras do tipo *Inkjet*, é possível inserir uma matriz 15 X 8 pontos;
- ▶ Essa matriz é feita de pontos amarelos, que são praticamente invisíveis ao olho humano, quando na luz ambiente;
- ▶ Essa matriz armazena informações da impressora de origem, como:
  - Data e Número de Série.



# Técnicas de Identificação

- ▶ Porém, existe uma maneira muito simples de burlar essa técnica;
- ▶ Basta imprimir uma folha em branco em uma impressora colorida, afim de obter a matriz dela;
- ▶ E imprimir o documento na mesma folha, mas em uma impressora monocromática, pois ela não é capaz de imprimir a matriz.

# Técnicas de Identificação

- ▶ A segunda técnica, faz um *shift* de 1/600 de polegada em algumas linhas do documento;
- ▶ Esse *shift* é característico de cada impressora.
- ▶ Dessa forma, não é possível fazer ataques do tipo “*scan-print*”;
- ▶ Mas, *softwares* de leitura de caracteres, podem quebrar essa técnica.

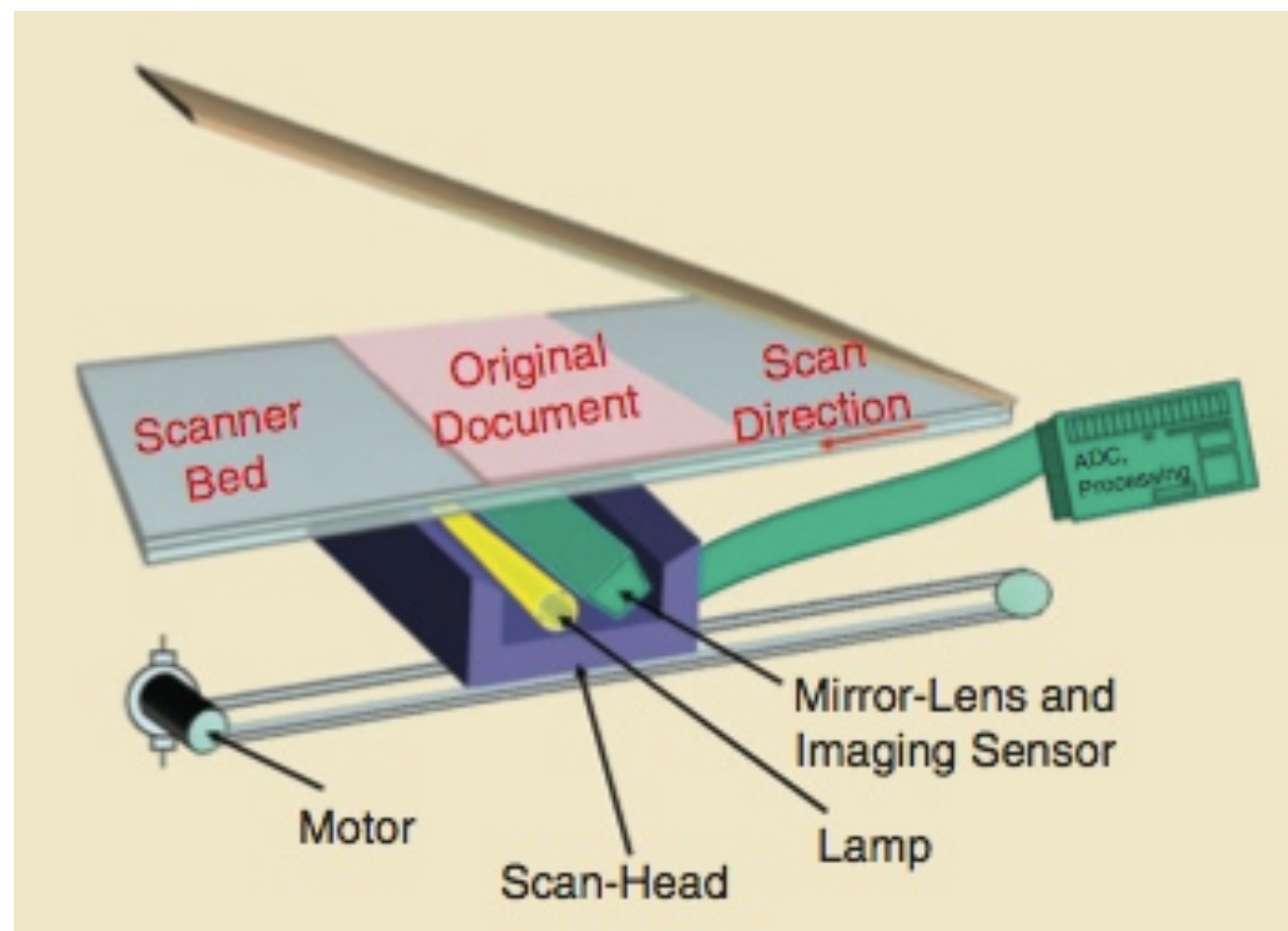
***Scanners***

# Scanners

- ▶ Três componentes principais
  - Luz;
  - Espelhos e sensores;
  - Motor de passo.
- ▶ Utilizar os defeitos desses componentes para atribuição.

# Scanners

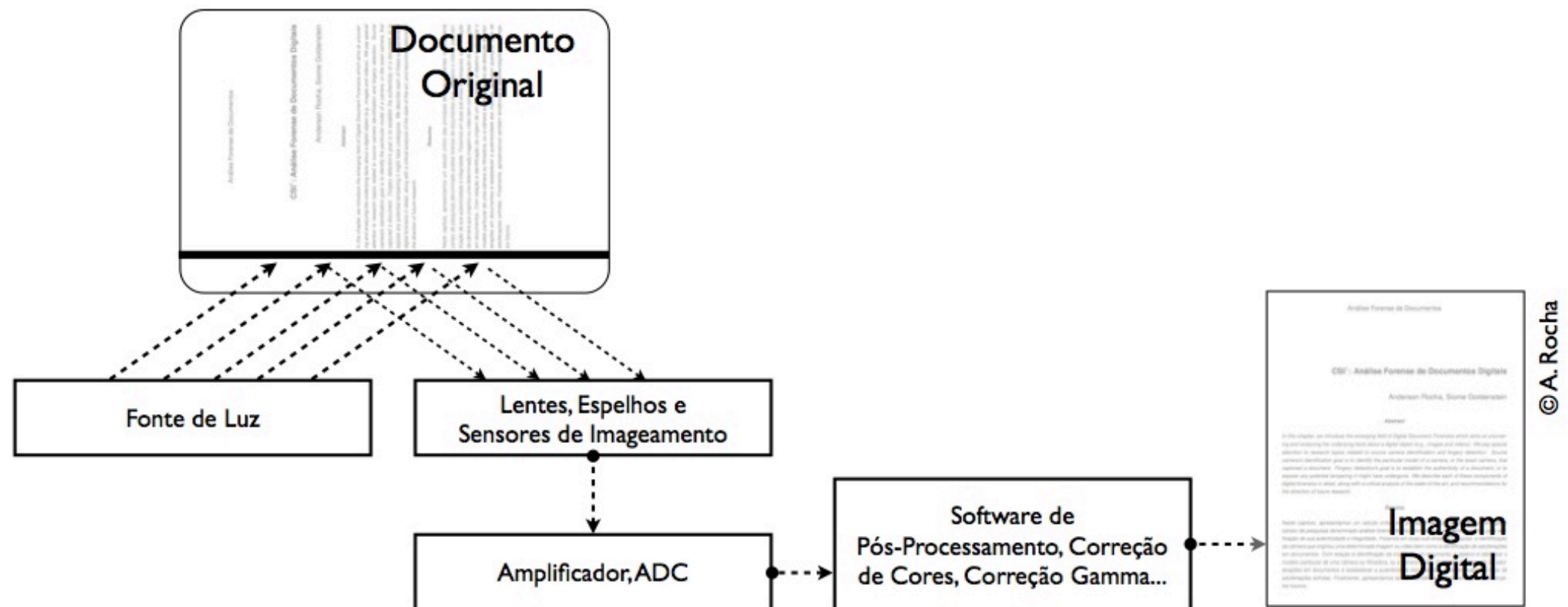
## ► Funcionamento



© P. Chiang

# Scanners

## ► Pipeline de um scanner



# Scanners

- ▶ A resolução nativa de um *scanner* é dada pela resolução do sensor (horizontal) e do motor de passo (vertical);
- ▶ Captura de imagens fora da resolução nativa, podem ser feitas de duas maneiras:
  - Sub-amostragem;
  - Amostragem normal, seguida de ajustes.

# Técnicas de Identificação

- ▶ Como obter a “assinatura” de um *scanner*?
- ▶ Similar as câmeras digitais, porém os resultados mostraram pouco acerto [Lukas et al. 2006];
- ▶ Três tipos de ruídos:
  - *Dead pixels, pixel trap*;
  - Ruído padrão;
  - Ruído aleatório.



# Técnicas de Identificação

## ► Filtro de redução de ruído

$$I_{noise}^k = I^k - I_{denoised}^k$$

## ► Com K imagens, podemos contruir um padrão de referência de um *scanner*

$$\tilde{I}_{noise}^{array}(i, j) = \frac{1}{K} \sum_{k=1}^K I_{noise}^k(i, j);$$

$$1 \leq i \leq M \text{ e } 1 \leq j \leq N$$

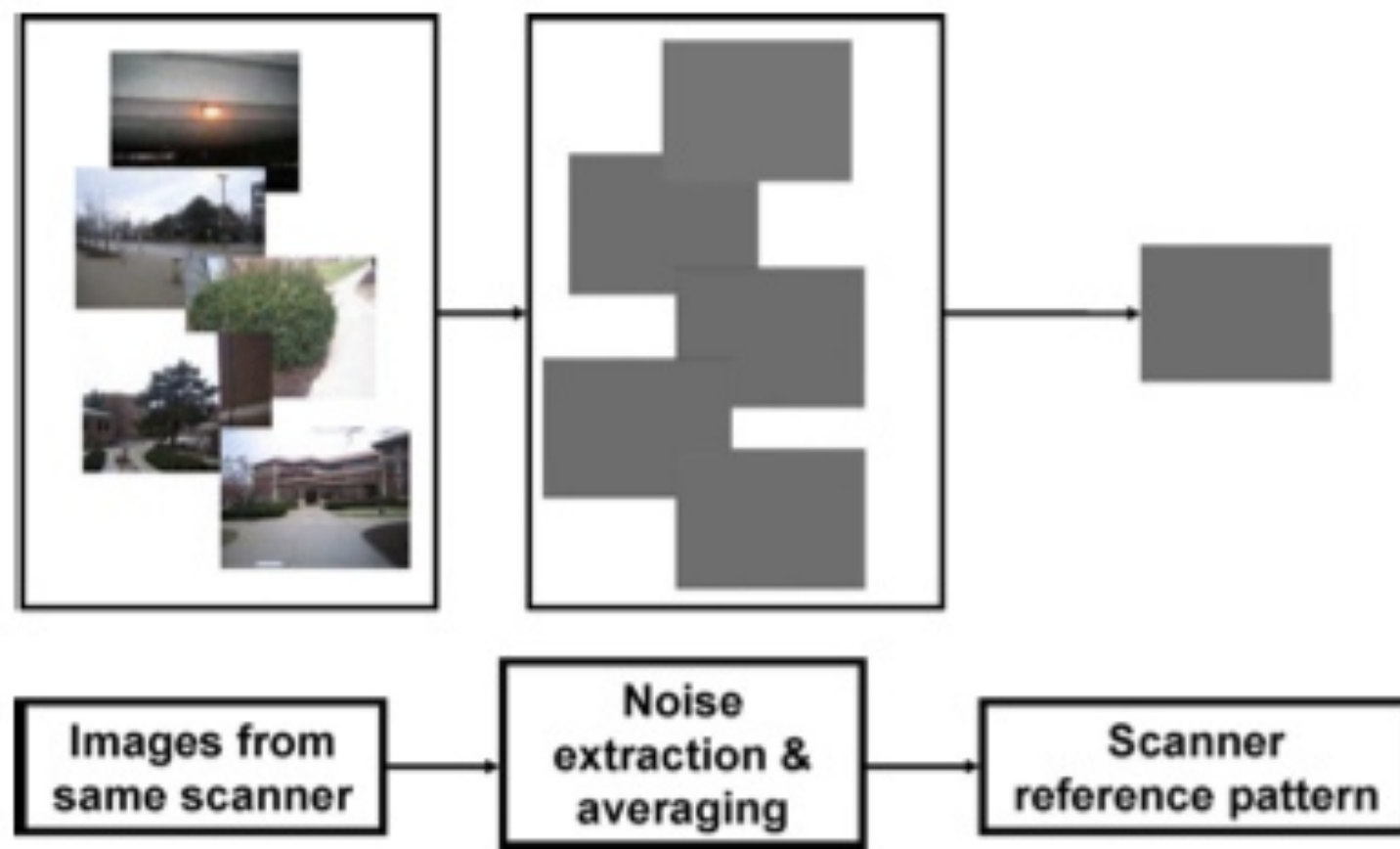
# Técnicas de Identificação

- Correlação entre um padrão de referência e a assinatura de um *scanner*

$$C(X, Y) = \frac{(X - \bar{X}) \cdot (Y - \bar{Y})}{||X - \bar{X}|| \cdot ||Y - \bar{Y}||}$$

# Técnicas de Identificação

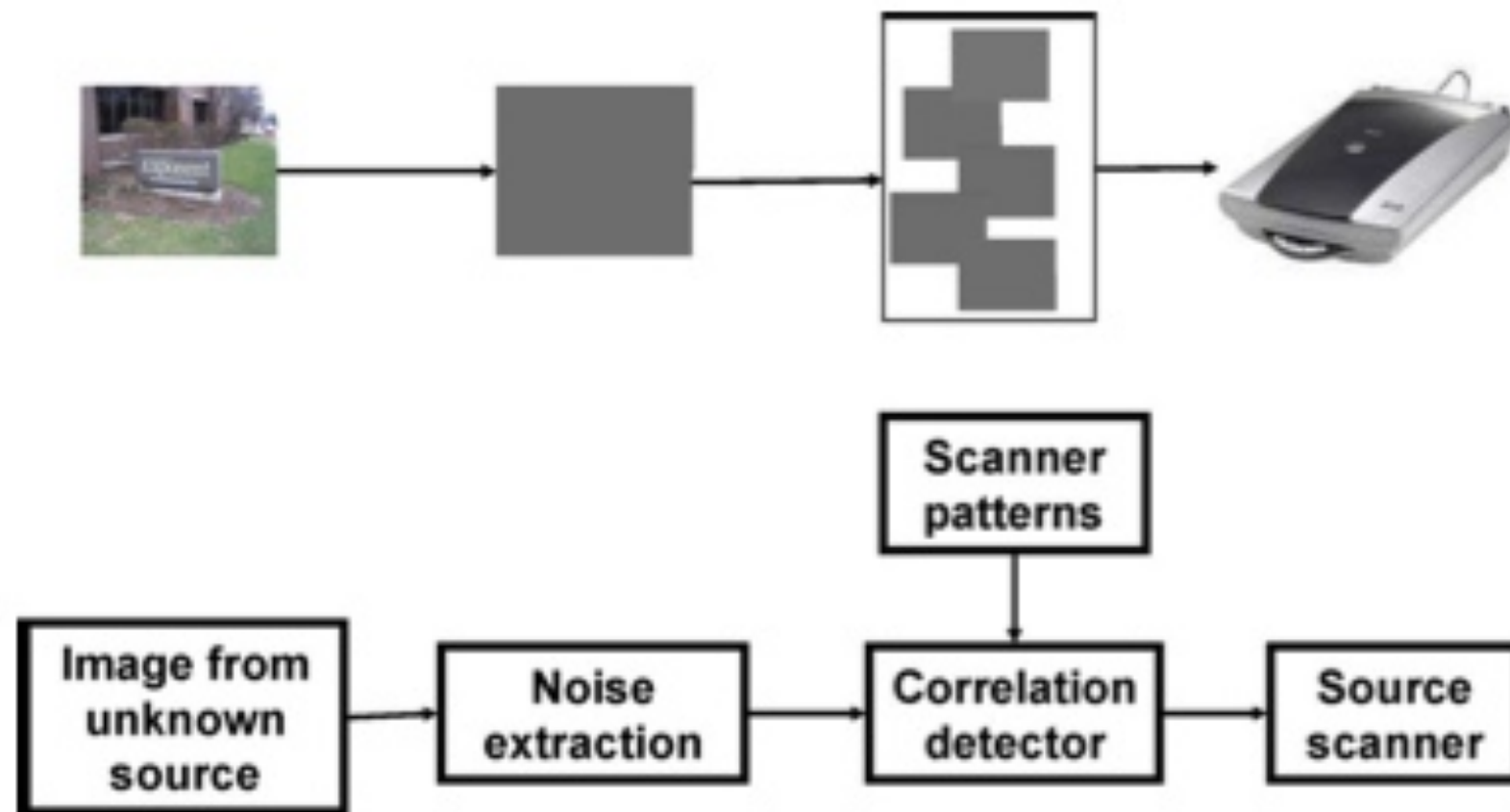
## ► Pipeline da técnica



© N. Khanna

# Técnicas de Identificação

## ► Pipeline da técnica



© N. Khanna

# Técnicas de Identificação

- Média dos ruídos das linhas e colunas

$$I_{\text{noise}} = I - I_{\text{denoised}}$$

$$\tilde{I}_{\text{noise}}^r(1, j) = \frac{1}{M} \sum_{i=1}^M I_{\text{noise}}(i, j); \quad 1 \leq j \leq N$$

$$\tilde{I}_{\text{noise}}^c(i, 1) = \frac{1}{N} \sum_{j=1}^N I_{\text{noise}}(i, j); \quad 1 \leq i \leq M.$$

# Técnicas de Identificação

- ▶ Correlação das linhas e colunas

$$\rho_{\text{row}}(i) = \mathbf{C} \left( \tilde{I}_{\text{noise}}^r, I_{\text{noise}}(i, \cdot) \right)$$
$$\rho_{\text{col}}(j) = \mathbf{C} \left( \tilde{I}_{\text{noise}}^c, I_{\text{noise}}(\cdot, j) \right) .$$

- ▶ É esperado que  $\rho_{\text{row}}$  seja maior que  $\rho_{\text{col}}$  por causa da periodicidade do sensor;

# Técnicas de Identificação

- ▶ A média, o desvio padrão, assimetria e a curtose de  $\rho_{row}$  e  $\rho_{col}$  são os 8 primeiros descritores;
- ▶ O desvio padrão, assimetria e a curtose de  $I_{noise, row}$  e  $I_{noise, col}$  formam 14 descritores;
- ▶ O último descritor é dado por:

$$f_{15} = \left( 1 - \frac{\frac{1}{N} \sum_{j=1}^N \rho_{col}(j)}{\frac{1}{M} \sum_{i=1}^M \rho_{row}(i)} \right) * 100$$

# Técnicas de Identificação

- ▶ Ao final de vários filtros, o autor chega em um vetor de 204-D, descritores;
- ▶ Classificador SVM com vetor de 10-D;
  - As 10-D, são combinações lineares das 204-D iniciais;



# Resultados

## ► Scanners utilizados

	Marca / Modelo	Sensor	Resolução Nativa
S <sub>1</sub>	Epson Perfetion 4490-Photo	CCD	4800 DPI
S <sub>2</sub>	HP ScanJet 6300c-1	CCD	1200 DPI
S <sub>3</sub>	HP ScanJet 6300c-2	CCD	1200 DPI
S <sub>4</sub>	HP ScanJet 8250	CCD	4800 DPI
S <sub>5</sub>	Mustek 1200 III Ep	CCD	1200 DPI
S <sub>6</sub>	Visoneer OneTouch 7300	CIS	1200 DPI
S <sub>7</sub>	Canon LiDE 25	CIS	1200 DPI
S <sub>8</sub>	Canon LiDE 70	CIS	1200 DPI
S <sub>9</sub>	OpticSlim 2420	CIS	1200 DPI
S <sub>10</sub>	Visoneer OneTouch 7100	CCD	1200 DPI
S <sub>11</sub>	Mustek ScanExpress A3	CCD	600 DPI

# Resultados

- ▶ Método de seleção dos *scanners* e imagens:
  - $S_1, S_2, S_3, S_4, S_6, S_7$  e  $S_9$ ;
  - 40 Imagens por *scanner*;
  - Cortadas em imagens de 1024x768;
  - Total de 200 imagens por *scanner*.

# Resultados

## ► Corte

B0	B1	B2	B3	B4
B5	B6	B7	B8	B9
...	...			

# Resultados

## ► Exemplos de Imagens



© N. Khanna

# Resultados

## ► Subimagens - Classes separadas JPEG (Q=70)

		Predicted													
		$S_1^1$	$S_1^2$	$S_2^1$	$S_2^2$	$S_3^1$	$S_3^2$	$S_4^1$	$S_4^2$	$S_6^1$	$S_6^2$	$S_7^1$	$S_7^2$	$S_9^1$	$S_9^2$
Actual	$S_1^1$	88.8	10.3	0	0.3	0.1	0.0	0.1	0	0	0	0.1	0.3	0	0.0
	$S_1^2$	22.8	76.1	0	0	0.0	0.3	0	0	0	0	0.0	0.7	0	0
	$S_2^1$	0	0	64.8	24.9	5.5	3.6	0.0	0.1	0.2	0	0.8	0	0.1	0
	$S_2^2$	0.3	0	20.2	57.9	7.4	12.2	0.5	0	0.2	0	0.3	0.5	0.5	0
	$S_3^1$	0.7	0.1	2.2	7.2	60.0	26.5	0	0.3	1.3	0.5	0.0	0	1.0	0.2
	$S_3^2$	0.4	0.0	4.9	11.2	18.7	60.1	0.0	0	0.2	0	1.3	0.5	1.7	0.8
	$S_4^1$	0.5	0	0	0	0	0	88.5	11.0	0	0	0	0	0	0
	$S_4^2$	0	0	0.0	0	0.1	0	8.9	91.0	0	0	0	0	0	0
	$S_6^1$	0	0	0	0.1	1.3	0.2	0	0	73.2	20.2	0.3	0	3.8	0.9
	$S_6^2$	0	0	0	0.1	0.2	0.3	0	0	31.6	64.5	0.8	0.3	1.4	0.9
	$S_7^1$	0	0	0.7	0.1	0.2	1.2	0	0	0.1	0	72.2	22.0	1.5	2.1
	$S_7^2$	1.0	0.7	0	0.3	0.2	2.0	0	0	0	0	20.0	71.9	0.3	3.4
	$S_9^1$	0	0	0.0	0.1	0.7	0.9	0.0	0	3.5	1.9	1.0	0.5	72.3	19.1
	$S_9^2$	0	0	0	0	0.3	1.2	0	0	2.0	1.7	1.8	2.1	23.0	67.9

# Resultados

## ► Subimagens - Classe única JPEG (Q=70)

		Predicted						
		$S_1$	$S_2$	$S_3$	$S_4$	$S_6$	$S_7$	$S_9$
Actual	$S_1$	99.2	0.2	0.2	0.0	0	0.4	0
	$S_2$	0.1	83.3	15.6	0.3	0.0	0.5	0.1
	$S_3$	0.7	10.1	85.3	0.2	1.4	0.8	1.6
	$S_4$	0.5	0	0	99.5	0	0	0
	$S_6$	0	0.3	0.9	0	93.9	0.6	4.2
	$S_7$	1.1	0.7	1.1	0	0.0	92.9	4.2
	$S_9$	0	0	1.8	0	5.2	3.0	90.0



# Resultados

## ► Subimagens - Classe única TIFF (**Resolução não nativa**)

		Predicted										
		$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$	$S_7$	$S_8$	$S_9$	$S_{10}$	$S_{11}$
Actual	$S_1$	100	0	0	0	0	0	0	0	0	0	0
	$S_2$	0	93.2	6.8	0	0	0	0	0	0	0	0
	$S_3$	0	10.1	89.9	0	0	0	0	0	0	0	0
	$S_4$	0	0	0	100	0	0	0	0	0	0	0
	$S_5$	0	0	0	0	100	0	0	0	0	0	0
	$S_6$	0	0	0	0	0	100	0	0	0	0	0
	$S_7$	0	0	0	0	0	0	100	0	0	0	0
	$S_8$	0	0	0	0	0	0	0	100	0	0	0
	$S_9$	0	0	0	0	0	0	0	0	100	0	0
	$S_{10}$	0.3	0	0	0	0	0	0	0.2	0	99.5	0
	$S_{11}$	0	0	0	0	0	0	0	0	0	0	100

# Conclusões



# Conclusão

- ▶ Modelo para impressoras se baseia na maioria das vezes nos defeitos mecânicos do equipamento;
- ▶ É possível distinguir entre dois scanners da mesma marca/modelo;
- ▶ Método é robusto tanto para imagens scaneadas em resolução nativa quanto não nativa;

# Referências

# Referências

1. **Bulan, Orhan, Junwen Mao e Gaurav Sharma:** *Geometric distortion signatures for printer identification*. *Acoustics, Speech, and Signal Processing*, IEEE International Conference on, 0:1401–1404, 2009.
2. **Chiang, Pei ju, Gazi N. Ali, Aravind K. Mikkilineni, Edward J. Delp, Jan P. Allebach e George T. c. Chiu:** *Extrinsic signatures embedding using exposure modulation for information hiding and secure printing in electrophotographic devices*. In *Proceedings of the IS&T's NIP20: International Conference on Digital Printing Technologies*, páginas 295–300, 2004.
3. **Gou, H., A. Swaminathan e M. Wu:** *Robust scanner identification based on noise features*. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 6505 de *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, fevereiro 2007.
4. **Kee, Eric e Hany Farid:** *Printer profiling for forensics and ballistics*. In *MM&#38;Sec '08: Proceedings of the 10th ACM workshop on Multimedia and security*, páginas 3–10, New York, NY, USA, 2008. ACM, ISBN 978-1-60558-058-6.
5. **Khanna, N., A.K. Mikkilineni e E.J. Delp:** *Scanner Identification Using Feature-Based Processing and Analysis*. *Information Forensics and Security*, IEEE Transactions on, 4(1):123 –139, mar. 2009, ISSN 1556-6013.

---

***Obrigado!***

---