### Detecção de códigos QR em imagens com enquadramento arbitrário

Luiz Felipe Franco Belussi

Dissertação apresentada ao Instituto de Matemática e Estatística da Universidade de São Paulo para obtenção do título de Mestre em Ciências

Programa: Ciência da Computação Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Nina Sumiko Tomita Hirata

São Paulo, setembro de 2012

#### Detecção de códigos QR em imagens com enquadramento arbitrário

Esta versão da dissertação/tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 30/07/2012. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof<sup>a</sup>. Dr<sup>a</sup>. Nina Sumiko Tomita Hirata (orientadora) IME-USP
- Prof. Dr. Anderson de Rezende Rocha IC-UNICAMP
- Prof. Dr. Paulo André Vechiatto de Miranda IME-USP

## Agradecimentos

Em primeiro lugar agradeço a meus pais Luiz e Normeli que me incentivaram desde a infância a buscar o conhecimento e realizaram muitos sacrifícios em nome da educação de seus filhos e em favor da construção de uma família feliz, sólida e de muito respeito. Eles são espetaculares.

Agradeço à minha esposa e companheira Fernanda Luiza que compreendeu que o mestrado era um sonho para mim e, assim sendo, permaneceu reclusa em minha companhia durante muitos finais de semana ensolarados para tornar esse trabalho possível.

Lilian e Caio, meus irmãos queridos, inspiraram-me durante toda a jornada, ela com sua bela e dedicada trajetória acadêmica e ele com as palavras de um verdadeiro amigo, por isso os agradeço sinceramente.

Agradeço à professora Nina Hirata, minha orientadora de valor incalculável, por seu exemplo de excelência e humildade, bem como por ter acreditado em meu potencial quando bati à sua porta procurando uma oportunidade.

As contribuições dos professores Anderson de Rezende Rocha do IC-UNICAMP, Carlos Morimoto e Paulo Miranda do IME-USP ajudaram a dar a forma final do trabalho e por isso lhes sou muito grato.

Enfatizo meus agradecimentos à Petrobras, mais especificamente na figura de meu gerente e amigo Janio, que teve papel fundamental nessa conquista ao entender a importância dessa realização acadêmica em minha vida e oferecer seu apoio sempre que precisei.

Por fim agradeço a Deus, pela luz, paz e disposição durante esses anos ocupados e felizes de minha vida.

### Resumo

BELUSSI, L. F. F. **Detecção de códigos QR em imagens com enquadramento arbitrário**, 2012. 85 f. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2012.

A utilização de códigos QR por deficientes visuais e robôs expande o horizonte de emprego dessa tecnologia e viabiliza várias aplicações cogitadas por um número crescente de trabalhos acadêmicos. Para que os códigos QR possam ser decodificados eles precisam ser apropriadamente enquadrados. A detecção desses símbolos em imagens adquiridas ao acaso é a primeira etapa na construção de um sistema de enquadramento assistido. O presente trabalho apresenta uma proposta para a detecção de códigos QR em imagens adquiridas ao acaso. A abordagem proposta é baseada em componentes e dividida em estágios, nos quais primeiramente partes do símbolo são detectadas através da técnica de detecção rápida de objetos proposta por Viola-Jones e em seguida uma análise do conjunto formado por essas partes é realizada para determinar a localização exata do símbolo na cena. A adaptação da técnica proposta por Viola e Jones para esse novo domínio é descrita e detalhada. É realizada uma discussão sobre a influência dos muitos parâmetros de treinamento e detecção presentes na abordagem proposta com base em resultados experimentais detalhados. Por fim é apresentado o desempenho em termos da qualidade dos resultados e tempo de processamento em vídeo e conclui-se que a abordagem proposta aponta um caminho promissor na resolução do problema de detecção de códigos QR em imagens adquiridas arbitrariamente na maioria das situações práticas.

Palavras-chave: código QR, código de barras bidimensional, ondaletas de Haar.

### Abstract

BELUSSI, L. F. F. **QR Code detection in arbitrarily acquired images**, 2012. 85 p. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2012.

The use of QR codes by visually impaired people and robots expands the applications of this technology and enables many possibilities that are emerging in a crescent number of scientific papers. QR code decoding relies on prior framing of the symbol during image acquisition. The detection of these symbols is the first step in the construction of a system to assist the framing process. This work presents an approach for the detection of QR codes in arbitrarily acquired images. The proposed approach is a component-based one consisting of two stages, where parts of the symbol are first detected through the fast object detection method proposed by Viola-Jones, and then an analysis of the set of detected parts is carried subsequently in order to determine the exact location of the symbol in the scene. The adaptation of the framework proposed by Viola-Jones for this new application domain is described in detail. The influence of various training and detection parameters of the framework in QR code detection is discussed using a series of experiments. The performance of the proposed approach is presented in terms of result quality and processing time and it is concluded that the proposed approach indicates a promising direction in solving the problem of detecting QR codes in arbitrarily acquired images in the majority of real world situations.

Keywords: QR code, two-dimensional barcode, Haar wavelets.

# Sumário

Lista de Figuras iz				ix		
Lista de Tabelas			xi			
1	Introdução					
	1.1	Motiv	ação	2		
	1.2	Objet	vos e contribuições	2		
	1.3	Organ	ização do Texto	3		
<b>2</b>	Detecção de códigos QR					
	2.1	Códig	os de Barras Bidimensionais	5		
	2.2	Estrut	ura do código QR	6		
	2.3	Revisa	ío Bibliográfica	7		
	2.4	Requi	sitos de Detecção em Imagens com Enquadramento Arbitrário	10		
3	Abordagem Viola-Jones para Deteção Rápida de Objetos					
	3.1	Haar-	like features	11		
	3.2	Image	m Integral	12		
	3.3	Casca	ta de classificadores	14		
	3.4	Boosta	ing	15		
4	Abordagem Proposta para Deteção de códigos QR					
	4.1	Visão	Geral da Abordagem Proposta	19		
	4.2	Detec	ção de FIPs usando Viola-Jones	21		
	4.3	Agreg	ação de Candidatos a FIP para Detecção de Códigos QR	21		
		4.3.1	Formulação do Problema	21		
		4.3.2	Algoritmo de Agregação de FIPs	24		
5	Treinamento e avaliação de detector de FIPs					
	5.1	5.1 Implementação utilizada				
		5.1.1	Utilitários do OpenCV	27		
		5.1.2	Algoritmo de <i>Boosting</i>	28		
		5.1.3	Parâmetros de Treinamento	29		

		5.1.4	Avaliação do Detector de FIP	30
	5.2	Deter	minação dos parâmetros de treinamento	32
		5.2.1	Conjunto de Dados	32
		5.2.2	Determinação experimental de parâmetros de treinamento	33
		5.2.3	Aumento das Amostras de Treinamento	37
6	$\mathbf{Res}$	ultado	s Experimentais	39
	6.1	Conju	ntos de Dados	39
	6.2	Detec	ção de FIPs $\ldots$ $\ldots$ $\ldots$	40
		6.2.1	Parâmetros de detecção	40
	6.3	Detec	ção de Códigos QR	43
		6.3.1	Parâmetros de detecção	43
		6.3.2	Tolerância dos critérios de restrições geométricas	43
		6.3.3	Número mínimo de detecções sobrepostas em códigos QR completos .	44
	6.4	Avalia	ção do Detector de Códigos QR	46
		6.4.1	Avaliação Quantitativa	46
		6.4.2	Avaliação Qualitativa	51
	6.5	Proces	ssamento de Vídeo	62
		6.5.1	Tamanho Mínimo de Detecção de FIPs no Processamento de Vídeo $\ .$	62
		6.5.2	Fator de Escalonamento no Processamento de Vídeo	63
	6.6	Discus	55ão	64
7	Con	clusão	o e Trabalho Futuro	67
Re	e <b>ferê</b> :	ncias l	Bibliográficas	69

# Lista de Figuras

2.1	Exemplos de Códigos de Barras (wikipedia)	5
2.2	Esturura do código QR (FIP: Padrões de posição / TP: Padrões de sincroni-	
	zação / AP: Padrões de alinhamento).	6
2.3	Proporção entre pixels brancos e pretos nos FIPs em linhas de busca diagonais	
	(d), verticais (v) e horizontais (h)	7
2.4	Diferentes versões de símbolos código QR	8
2.5	Exemplo de marcador visual utilizado para facilitar a detecção de códigos ${\rm QR}$	
	usada em Xue <i>et al.</i> (2010)	9
3.1	Conjunto básico de protótipos de características	12
3.2	Conjunto ampliado de protótipos de características	12
3.3	Exemplos de características haar-like features.	13
3.4	A soma dos pixels da área sombreada é dada por: $ii(z) - ii(y) - ii(w) + ii(x)$ .	14
3.5	Cascata de classificadores.	14
4.1	Um código QR de versão 1 é composto por 21 $\times$ 21 módulos	22
4.2	Dados dois FIPs (indicados como 1 e 2) de um mesmo código QR, as seis	
	possíveis posições que um terceiro FIP pode ocupar são representadas pelos	
	círculos negros.	22
4.3	Uma aresta que pode ser parte de dois subconjuntos de três candidatos a FIP	
	satisfazendo os critérios.	24
5.1	Uma imagem de fundo com a inserção de um FIP	30
5.2	Escolha de parâmetros: em cada nível o nó de cor escura corresponde ao valor	
	escolhido.	34
5.3	Recuperação $\times$ FP para classificadores obtidos com diferentes parâmetros de	
	treinamento	35
5.4	Duas características principais para MODE=SYM (linha superior) e MODE=NO	ONSYM
	(linha inferior). $\ldots$	36
5.5	$\operatorname{Exemplos}$ de amostras geradas a partir de variações aleatórias de uma amostra	
	positiva.	37
5.6	Efeito do aumento de amostras positivas de treinamento, média de 5 rodadas.	38

6.1	Recuperação e FP apresentados por um mesmo detector de FIP com variação $\hfill \hfill \hfil$	
	do parâmetro de detecção SF (fator de escalonamento)	41
6.2	Recuperação $\times$ FP para o detector de FIPs treinado com variação do pa-	
	râmetro de detecção ND (número mínimo de detecções sobrepostas) de 1 a $\ $	
	200	42
6.3	Recuperação e FP do detector de códigos QR completos (estágios 1 e 2)	
	sujeitos à variação do parâmetro de tolerância do critério de geometria $d\ {\rm com}$	
	valor fixo para o parâmetro de tolerância do critério de tamanho $\varepsilon=0.30.$	44
6.4	Recuperação e FP apresentados pelo detector de códigos QR completos (está-	
	gios 1 e 2) enquanto é sujeito à variação do parâmetro de tolerância do critério	
	de tamanho $\varepsilon$ com valor fixo para o parâmetro de tolerância do critério de	
	geometria $d = 0.25$ .	45
6.5	Recuperação e FP do detector de códigos QR completos (estágios $1 e 2$ ) com	
	valores fixos dos parâmetros $\varepsilon=0.4$ e $d=0.25$ sujeitos à variação do número	
	mínimo de detecções sobrepostas (ND)	45
6.6	Múltiplos códigos QR próximos na imagem (www.flickr.com blech).	52
6.7	Duas detecções corretas e um falso positivo. (www.flickr.com, scott_blake)	53
6.8	Uma detecção correta e um falso negativo devido a um FIP não detectado. (www.	flickr.com
	, davidking) $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$	54
6.9	Uma detecção com oclusão parcial de um FIP e um falso negativo, possivel-	
	mente devido à inclinação. (www.flickr.com, cjsveningsson)	55
6.10	Três falsos negativos. (www.flickr.com davidking)	55
6.11	Uma detecção correta e outra formada por três falsos positivos para FIP. (www.fli	$\operatorname{ckr.com}$
	$, qriouscode) \ldots \ldots$	56
6.12	Exemplo de detecção correta de um código QR (www.flickr.com, gaku). $\ .$ .	57
6.13	Exemplo de detecção correta de um código QR com vários falsos positivos	
	para FIPs (www.flickr.com, whoisstan).	58
6.14	Exemplo de detecção correta de um código QR com vários falsos positivos	
	para FIPs (www.flickr.com, whoisstan).	59
6.15	Exemplo de detecção correta de um código QR (www.flickr.com, whoisstan).	60
6.16	Exemplo de detecção correta de um código QR com dois falsos positivos para $\hfill$	
	FIPs (www.flickr.com, whoisstan).	61
6.17	Quadro retirado de um dos fluxos de vídeo experimentais	63
6.18	Tempo de processamento dos estágios $1 \neq 2$ (resultados em ms, fluxo de vídeo	
	de 640 $\times$ 480 pixels, média de 1000 quadros)	64
6.19	Tempo de processamento para os estágios 1 e 2 (resultados em ms, fluxo de $\hfill$	
	vídeo de 640 × 480 pixels, valores médios para 1000 quadros). $\ldots$	65

# Lista de Tabelas

6.1	Resultados para a detecção de FIPs - Rodada 1	47
6.2	Resultados para a detecção de FIPs - Rodada 2	48
6.3	Resultados para a detecção de códigos QR completos - Rodada 1. $\ldots$	49
6.4	Resultados para a detecção de códigos QR completos - Rodada 2. $\ldots$	49
6.5	Resultados para a detecção de FIPs	50
6.6	Resultados para a detecção de códigos QR completos	51

#### xii LISTA DE TABELAS

## Capítulo 1

## Introdução

Códigos de barras têm como aplicação tradicional a identificação de objetos de forma simples e eficiente ao possibilitar que as informações codificadas em uma etiqueta possam ser facilmente lidas e interpretadas por um sistema de informação. Trata-se de uma tecnologia madura e extremamente bem sucedida, tendo sido adotada amplamente em diversos contextos, principalmente em publicidade e na identificação de produtos. Atualmente ela tem-se mostrado adequada a novos domínios de aplicação com a popularização de telefones celulares equipados com câmeras e o desenvolvimento de padrões de codificação de alta capacidade.

Esse trabalho aborda a detecção de códigos QR, um tipo de código de barras bidimensional que vem ocupando lugar de destaque entre outros padrões de mercado e demonstrando crescente popularidade no mundo todo. O código QR foi criado pela empresa Denso Wave em 1994, está descrito na ISO/IEC 18004 e seu nome vem do inglês "Quick Response Code". Uma variedade de aplicações tem sido considerada em âmbito comercial e científico sendo que algumas das possibilidades mais inovadoras se beneficiariam muito da existência de técnicas para reconhecimento automatizado desses códigos de barras em cenas capturadas ao acaso. Dentre as possibilidades de aplicação do reconhecimento automatizado de códigos QR destacam-se sistemas de acessibilidade para deficientes visuais e a construção de robôs autônomos que utilizem as informações presentes nesses códigos para cumprir suas tarefas.

Uma pesquisa na literatura especializada revelou que a detecção de códigos de barras tradicionalmente se refere à determinação da localização exata do símbolo em imagens adquiridas com o propósito específico de capturá-lo e que existe uma lacuna no tratamento de cenas com enquadramento arbitrário - um problema totalmente distinto e mais complexo.

O presente trabalho propõe uma abordagem para a detecção rápida de códigos QR em imagens com enquadramento arbitrário sem fazer uso de indicações auxiliares e que pode ser utilizada no processamento de vídeo em tempo real. O objetivo, além de detectar a presença desses códigos de barras nas imagens, é delimitar de maneira precisa sua posição em tais imagens pois, dessa forma, é possível orientar um usuário ou um robô na realização de um enquadramento apropriado do código.

#### 1.1 Motivação

Códigos QR foram concebidos para uso na indústria automotiva como forma de rastrear peças ao longo do processo de manufatura (DensoWave). Atualmente os códigos QR têm sido predominantemente usados como apontadores físicos (*hyperlinks* codificados) para conectar lugares e objetos do mundo físico a conteúdo na *internet* relacionado ao contexto em que se encontram. Estão surgindo, com base em cógidos QR, aplicações educacionais e de entretenimento (Ceipidor *et al*, 2009; Ozcelik e Acarturk 2011; Susono e Shimomura 2006), de segurança da informação e de sistemas (Chen e Wang 2009; Liao e Lee 2010; Liu 2010), para o oferecimento de serviços específicos (Canadi *et al*, 2010; Walsh 2010) além de toda sorte de aplicações relacionadas a dispositivos móveis (*tablets* e *smartphones*). Academicamente temse vislumbrado a possibilidade de utilização de códigos QR em aplicações mais elaboradas para auxílio a deficientes visuais - parciais ou totais - ou mesmo robôs autônomos na conclusão de tarefas como navegação interna em prédios, compras, leitura e outras (Al-Khalifa 2008; Coughlan *et al*, 2006; Ebrahim *et al*, 2005).

Os detectores de códigos QR existentes atualmente, comumente encontrados em dispositivos móveis, dependem do correto enquadramento do símbolo a ser detectado que deve ocupar aproximadamente, no mínimo, 30% da imagem capturada. Ao explorar um ambiente, deficientes visuais ou robôs em regra não são capazes de capturar imagens com tal enquadramento a menos que lhes seja indicada a localização do símbolo a detectar. Dessa maneira, para viabilizar a utilização das aplicações baseadas em códigos QR nos casos descritos, determinar a existência e localização de um código QR em uma imagem é um passo requerido para propiciar o correto enquadramento do símbolo e deve anteceder o processo de decodificação.

Na literatura relacionada a maioria dos trabalhos que mencionam o reconhecimento ou detecção de códigos QR está orientada a melhorar a qualidade de uma imagem para decodificação ou a determinar os contornos exatos de um símbolo presumidamente presente e bem enquadrado nessa imagem ao invés de encontrar (decidindo se há ou não e em que posição) um código QR em tal imagem (Chu *et al*, 2007; Liu *et al*, 2008b; Ohbuchi *et al*, 2004).

Alguns poucos artigos abordam o problema de detectar a presença e determinar a posição de códigos QR em imagens com enquadramento arbitrário e propõem soluções baseadas em informação auxiliar como marcações visuais ou, alternativamente, etiquetas de radiofrequência (RFID) (Coughlan *et al*, 2006; Xue *et al*, 2010). Apesar de esse tipo de abordagem ser viável em ambientes restritos e controlados, a necessidade de inserção de marcações artificiais no ambiente inviabiliza sua utilização em situações gerais.

#### 1.2 Objetivos e contribuições

Esse trabalho visa desenvolver uma abordagem técnica que possibilite a determinação de tamanho e posição de um código de barras bidimensional no padrão código QR em cenas

adquiridas ao acaso (i.e. sem a intenção explícita de enquadrar o símbolo). Além disso a minimização do custo computacional dos algoritmos utilizados deve ser priorizada pois uma solução eficiente, que possibilite processamento em tempo real, tem potencial de aplicação consideravelmente mais amplo.

O trabalho propõe uma abordagem eficiente e eficaz para detecção de códigos QR em imagens adquiridas naturalmente, um problema para o qual não foram encontradas soluções na literatura pesquisada. A abordagem proposta é formulada como um problema de detecção baseada em componentes, seguida de agregação dos componentes detectados que são representados por partes da estrutura do código denominadas "FIPs" (do inglês "Finder Patterns"). Para a detecção de componentes, é proposto um inovador uso da técnica de detecção rápida de objetos de Viola e Jones, originalmente usada em detecção de faces. Para a etapa de agregação é proposta uma formulação baseada em grafos e coerência espacial e geométrica dos componentes detectados. São descritos também vários experimentos realizados tanto para o ajuste como para a avaliação dos detectores. Portanto, além de constituir uma contribuição concreta para o problema de detecção de códigos QR, este trabalho pode contribuir como referência para problemas que visam a detecção de outros tipos de objetos.

#### 1.3 Organização do Texto

No capítulo 2 são apresentados diferentes tipos de códigos de barra incluindo os bidimensionais e a estrutura do código QR é detalhada. O problema de detecção de códigos de barra em situação de enquadramento arbitrário é introduzido e uma revisão bibliográfica acerca do tema é feita.

O capítulo 3 expõe conceitos como *haar-like features*, cascata de classificadores, *boosting* e imagem integral, que são empregados no arcabouço para detecção de objetos proposto por Viola-Jones, utilizado nesse trabalho.

Em seguida, no capítulo 4 é descrita em detalhes a abordagem proposta para a solução do problema de detecção de códigos QR em cenas com enquadramento arbitrário. São feitas considerações sobre os conjuntos de treinamento utilizados, os diferentes parâmetros de treinamento da cascata de classificadores e a etapa de pós-processamento é descrita em detalhes.

A descrição das ferramentas utilizadas é feita no capítulo 5 que também apresenta os experimentos conduzidos para a determinação dos valores adequados dos parâmetros de treinamento do detector de FIPs.

No capítulo 6 são apresentados os resultados dos experimentos realizados para a determinação de parâmetros de detecção adequados ao problema estudado bem como os resultados quantitativos e qualitativos alcançados pela abordagem proposta. O tempo de processamento em fluxos de vídeo é apresentado e são fornecidos exemplos de detecções que evidenciam a qualidade dos resultados obtidos bem como os desafios ainda existentes. Parte dos resultados

#### 4 INTRODUÇÃO

apresentados deu origem a dois artigos, a saber:

- "Fast QR code detection in arbitrary acquired images" Belussi e Hirata (2011) publicado no XXIV SIBGRAPI.
- "Fast component-based QR code detection in arbitrarily acquired images" Belussi e Hirata (2012) publicado no Journal of Mathematical Imaging and Vision.

Finalmente em 7 são discutidos os resultados alcançados e apontadas as perspectivas de trabalho futuro.

## Capítulo 2

## Detecção de códigos QR

Este capítulo apresenta códigos de barras bidimensionais e contextualiza o problema de sua detecção em imagens arbitrárias. A estrutura dos códigos QR é apresentada em detalhes e em seguida é feita uma revisão bibliográfica que ressalta a demanda existente por uma solução ao problema de detecção de códigos QR em imagens adquiridas arbitrariamente.

#### 2.1 Códigos de Barras Bidimensionais

Os códigos de barras podem ser classificados quanto a sua dimensionalidade em 1D ou 2D, sendo que em alguns casos o uso de cores pode ser considerado uma dimensão adicional. Códigos de barras 2D apresentam uma densidade de informação muito maior que seus correspondentes unidimensionais e possibilitam o desenvolvimento de aplicações mais complexas. A Figura 2.1 apresenta exemplos de vários padrões de codificação em barras:



Figura 2.1: Exemplos de Códigos de Barras (wikipedia)

O código QR (ISO/IEC 18004) é o padrão de código de barras bidimensional que vem

recebendo maior atenção do mercado nos últimos anos, tornando-se referência no uso comercial desse tipo de tecnologia. Dentre os usos mais comuns para os códigos QR está a codificação de apontadores que possibilita aos utilizadores, ao apontar uma câmera para o código, estabelecer uma conexão entre objetos no mundo real (físico) e conteúdo na *internet*. A popularidade dos códigos QR pode ser creditada à robustez de decodificação, à alta capacidade de armazenamento e à rica especificação desse padrão de código de barras. Tal robustez tem sua importância evidenciada no próprio nome do padrão pois QR é uma abreviação de "Quick Response", uma alusão à detecção e decodificação rápidas mesmo em condições adversas.

#### 2.2 Estrutura do código QR

Segundo o padrão ISO/IEC 18004, a estrutura dos códigos QR contempla, além dos dados, informação de versão e correção de erros, os seguintes elementos estruturais<sup>1</sup> destacados na Figura 2.2:



**Figura 2.2:** Esturura do código QR (FIP: Padrões de posição / TP: Padrões de sincronização / AP: Padrões de alinhamento).

- Margem Silenciosa (em torno do símbolo)
- 3 Padrões de posição (nos cantos) FIP
- 2 Padrões de sincronização (entre os padrões de posição) TP
- N Padrões de alinhamento (distribuídos pela área do código) AP

 $<sup>^1\</sup>mathrm{A}$  nomenclatura das estruturas foi mantida em inglês uma vez que ainda não existem traduções para o português.

A margem silenciosa é uma margem em torno do código de barras que deve ser respeitada para facilitar a detecção do símbolo e deve ter ao menos 4 módulos de largura, sendo que cada módulo corresponde à menor unidade de informação no código (o menor quadrado preto ou branco).

Os *padrões de alinhamento* são usados para corrigir distorções não lineares ao longo da área do símbolo e têm maior importância quanto mais informação for codificada.

Os *Padrões de sincronização* indicam a localização das linhas e colunas de módulos no símbolo.

Os *padrões de posição* são usados para auxiliar na localização do símbolo e foram especialmente projetados para serem encontrados em qualquer direção uma vez que a proporção de alternância entre pixels brancos e pretos (na razão de 1:1:1:3:1:1) se mantém constante ao longo de linhas de escaneamento que passem através de seu centro, conforme ilustrado na Figura 2.3.



**Figura 2.3:** Proporção entre pixels brancos e pretos nos FIPs em linhas de busca diagonais (d), verticais (v) e horizontais (h).

De acordo com a quantidade de dados a ser codificada em um código QR, versões com capacidades distintas são utilizadas e as proporções existentes entre os tamanhos dos elementos estruturais do código e sua área total são alteradas. Portanto nota-se que códigos QR não podem ser considerados formas *rígidas* pois diferentes instâncias podem exibir diferentes proporções entre seus elementos estruturais e a área total do símbolo.

A Figura 2.4 ilustra através de exemplos a relação entre o número de módulos presentes em um símbolo e caracteres codificados. Além da diferença de tamanho, observam-se variações estruturais como variação no número de *padrões de alinhamento* e a mudança na relação entre o tamanho dos FIPs e a área total do código.

#### 2.3 Revisão Bibliográfica

A detecção de códigos de barras 2D da maneira como é descrita na literatura especializada e nas soluções de mercado refere-se à determinação de posição e tamanho de uma "caixa delimitadora" (*bounding box*) para a região de código em uma imagem adquirida com o propósito específico de evidenciá-la e destacá-la. Como consequência, os algoritmos usados



Figura 2.4: Diferentes versões de símbolos código QR

nessa aplicação apenas são capazes de detectar símbolos que ocupam uma parte representativa (ao menos 30%) da área enquadrada, com alguma tolerância a rotação e distorção perspectiva.

Em contrapartida, a detecção de códigos de barras 2D em imagens com enquadramento arbitrário representa um desafio maior pois, além de buscar símbolos menores (até 2% da área enquadrada), é necessário que seja realizada com velocidade muito maior que o processo de detecção convencional para que exista aplicação prática da técnica desenvolvida.

A importância do problema de detecção de códigos de barras 2D em imagens com enquadramento arbitrário é apontada reiteradas vezes na literatura, como em Chu *et al.* (2007) que, além de destacar o problema, menciona a ausência de trabalhos que o abordem.

Apesar de ser um problema relacionado a uma série de possíveis aplicações como ferramentas de apoio a deficientes visuais (Al-Khalifa 2008; Coughlan *et al*, 2006; Ebrahim *et al*, 2005) e processamento de código em vídeo (Liu *et al*, 2008a; Ohbuchi *et al*, 2004; Pârvu e Bălan , 2009), não foram encontradas referências específicas que tratem da detecção de códigos QR em imagens obtidas através de enquadramento arbitrário.

Em Ebrahim *et al.* (2005) os autores propõem uma abordagem para a detecção de códigos de barra 2D no padrão PDF417 com o intuito de ajudar deficientes visuais a identificar objetos por meio de tais códigos. Esse trabalho reconhece a dificuldade de se tratar cenas

com enquadramento arbitrário apontando como um dos maiores desafios o desenvolvimento de algoritmos eficientes para reconhecimento de códigos de barras em situações de pouca iluminação e em imagens de baixa resolução.

Alguns trabalhos como Al-Khalifa (2008) discutem as possibilidades de utilização de códigos QR para auxiliar na identificação de objetos etiquetados por deficientes visuais totais e parciais mas, apesar disso, não mencionam o problema de enquadramento por parte de uma pessoa nessas condições. Deficientes visuais têm pouca ou nenhuma chance de capturar imagens com um bom enquadramento do código e, dessa maneira, o problema de detecção dos códigos em imagens adquiridas com enquadramento arbitrário é base técnica para viabilizar aplicações como a proposta no referido artigo.

O problema pode ser abordado com o auxílio de etiquetas de rádiofrequência (RFID) ou, alternativamente, através do uso de marcadores visuais, como é o caso de Coughlan *et al.* (2006) e Xue *et al.* (2010). Um exemplo de emprego de marcador visual pode ser visto na Figura 2.5.



**Figura 2.5:** Exemplo de marcador visual utilizado para facilitar a detecção de códigos QR usada em Xue et al. (2010).

A abordagem de Xue *et al.* (2010) foi a única tentativa bem sucedida encontrada na literatura para a detecção de códigos QR em cenas com enquadramento arbitrário. Contudo essa abordagem tem uma séria limitação por se utilizar de marcadores visuais artificiais que acabam por inviabilizar seu uso em ambientes que não foram previamente preparados.

Uma solução que possibilite a detecção de códigos de barras 2D em imagens com enquadramento arbitrário de maneira computacionalmente eficiente e que dispensasse o uso de marcadores artificiais viabilizaria a construção de aplicações nas áreas de acessibilidade e robótica.

### 2.4 Requisitos de Detecção em Imagens com Enquadramento Arbitrário

O requisito de detecção mais importante ao levar em consideração aplicações em tempo real é que o processamento possa acontecer de maneira rápida, computacionalmente eficiente. Além disso é imprescindível que o detector seja invariante em escala uma vez que não há nenhum conhecimento prévio do tamanho em que se apresentará um símbolo eventualmente capturado na imagem a ser processada. Por fim, como não são consideradas restrições para a aquisição das imagens a serem processadas, idealmente o detector deve ser robusto a variações de iluminação, rotação e distorção perspectiva.

A indústria de processadores para dispositivos móveis tem apontado uma clara tendência ao aumento do número de núcleos. Assim sendo, uma abordagem paralelizável para a detecção de códigos QR tem chances aumentadas de atingir a velocidade de processamento adequada ao processamento de vídeo em tempo real nesse tipo de dispositivo.

Com o intuito de estabelecer limites à tarefa de detecção, neste trabalho considera-se como objeto de estudo as situações em que os códigos QR ocorrem com grande variação de escala e de iluminação mas com níveis reduzidos de rotação e distorção perspectiva. Há que se notar que códigos QR tipicamente são posicionados de forma alinhada com o chão os as bordas de uma publicação. Dessa forma é provável que, ao fotografar uma cena que contenha um código QR, o símbolo encontre-se pouco rotacionado.

Além disso é perfeitamente plausível considerar a detecção apenas de códigos QR escuros com fundo claro pois, apesar de a inversão dos tons dos códigos ser possível e estar prevista no padrão que os especifica, a ocorrência de símbolos claros com fundo escuro é muito baixa. Por outro lado, caso em determinado domínio de aplicação a ocorrência de códigos QR claros em fundo escuro seja alta, basta que seja tomado o complemento da imagem antes do processamento e o problema se torna o mesmo.

### Capítulo 3

# Abordagem Viola-Jones para Deteção Rápida de Objetos

Este capítulo resume conceitos técnicos empregados no arcabouço de detecção rápida de objetos desenvolvido por Viola e Jones (Viola e Joneș 2001), descrevendo-os e relacionando-os com os objetivos do trabalho. Viola e Jones desenvolveram uma abordagem chamada de *Boosted Cascade of Simple Features* que faz uso de uma árvore CART (Breiman *et al*, 1984), de características simples inspiradas em *haar wavelets* denominadas *haar-like features* de Papageorgiou e Poggio (2000) e da técnica de *boosting* (Freund e Schapire, 1997) para agregação de classificadores fracos.

Serão expostos os conceitos de *haar-like features*, de imagem integral, de cascateamento e *boosting* de classificadores.

#### 3.1 Haar-like features

O uso de características possibilita identificar estruturas gerais do padrão buscado de maneira muito mais eficiente do que o próprio valor dos pixels da imagem. A escolha do conjunto de características adequadas bem como a complexidade de seu cálculo representam desafios recorrentes no reconhecimento de padrões. Viola e Jones justificam o uso de características argumentando que através delas é possível operar em um domínio mais amplo e detectar cantos, bordas e linhas dentro da janela de detecção.

Nesse trabalho considera-se imagens em tons de cinza que representam a intensidade de luz presente na imagem capturada, ou seja, nessas imagens existe apenas um canal. Qualquer imagem colorida contendo múltiplos canais para representar diferentes componentes de cores pode ser convertida para a escala de cinza através da soma ponderada de suas componentes.

A técnica a ser apresentada utiliza características inspiradas em *Haar wavelets* que são em geral formas de ondas quadradas com diferentes períodos. Tais wavelets são usadas na transformada de Haar para possibilitar a decomposição de uma função em uma soma dessas wavelets multiplicadas por seus respectivos coeficientes, determinados durante o cálculo da mesma.

Viola e Jones propuseram o uso de características baseadas em Haar wavelets, cujos protótipos estão ilustrados na Figura 3.1, para o reconhecimento de padrões em imagens (Viola e Joneș 2001).



Figura 3.1: Conjunto básico de protótipos de características.

Posteriormente, Lienhart e Maydt (2002) estenderam o trabalho de Viola e Jones (2001) acrescentando ao conjunto original outros protótipos de características rotacionados a 45°, resultando no conjunto representado na Figura 3.2, utilizado nesse trabalho.



Figura 3.2: Conjunto ampliado de protótipos de características.

Uma característica é, nesse contexto, uma instância de um protótipo quando se define para esse sua posição relativa a uma janela de processamento, sua largura e sua altura. A Figura 3.3 mostra exemplos dessas características.

Para cada região da imagem sob a janela de processamento o valor de uma característica é calculado pela soma dos pixels contidos na área escura subtraídos da soma dos pixels contidos na área clara. Conforme mencionado anteriormente, a complexidade na avaliação do valor das características escolhidas tem papel crucial no desempenho do classificador que as utilizará. A imagem integral (descrita na seção 3.2) foi proposta como uma poderosa ferramenta na otimização do cálculo dessas características.

#### 3.2 Imagem Integral

Para que o cálculo das *haar-like features* possa ser realizado em múltiplas escalas com complexidade computacional razoável, o uso de uma imagem integral foi proposto por Viola e Jones (2001). A imagem integral, calculada como passo preliminar no algoritmo de detecção, é uma matriz com as mesmas dimensões da imagem de entrada na qual cada posição (x, y) armazena a soma dos valores de todos os pixels no retângulo compreendido entre (0,0) e (x, y). Dada uma imagem *i* de dimensão  $h \times w$ , o cálculo da imagem integral correspondente *ii* pode ser feito em aproximadamente  $2 \times (w \times h)$  acessos a memória com base na seguinte recorrência:



Figura 3.3: Exemplos de características haar-like features.

$$s(x,y) = s(x,y-1) + i(x,y)$$
(3.1)

$$ii(x,y) = ii(x-1,y) + s(x,y)$$
(3.2)

onde s(x, y) é a soma acumulada na linha, sendo que s(x, -1) = 0, e ii(-1, y) = 0.

Uma vez calculada a imagem integral, a soma dos valores dos pixels de qualquer retângulo contido na imagem pode ser determinada em exatamente 4 acessos à memória, independentemente de seu tamanho, conforme ilustrado na Figura 3.4. Dessa forma, qualquer característica como as apresentadas na seção 3.1 pode ser calculada em tempo constante.

Para o conjunto estendido de protótipos de características, o mesmo desempenho pode ser alcançado com uma variação da imagem integral introduzida por Lienhart e Maydt (2002) que utiliza triângulos ao invés de retângulos.

Como o cálculo de *haar-like features* corresponde a fazer a soma ou diferença da soma das intensidades em retângulos da imagem de interesse, o uso de imagem integral tem um impacto extremamente relevante na complexidade do algoritmo de detecção.



**Figura 3.4:** A soma dos pixels da área sombreada é dada por: ii(z) - ii(y) - ii(w) + ii(x).

#### **3.3** Cascata de classificadores

O arcabouço de Viola e Jones para detecção rápida de objetos, utilizado nesse trabalho, faz uso de uma cascata de classificadores construídos através da técnica de *boosting* (detalhada na seção 3.4). A composição em cascata apresenta a vantagem de direcionar o esforço computacional de busca para as áreas mais prováveis de ocorrência do padrão desejado. O emprego de *boosting* permite combinar o poder de múltiplos classificadores fracos baseados em *haar-like features*.

Uma classificadores entendida cascata de árvore pode ser como uma CART (Breiman et al, 1984) degenerada na qual cada estágio submete a amostra recebida de seu predecessor a um classificador e, de acordo com a saída desse, o processamento é encerrado ou continua em um estágio sucessor. Um encadeamento de estágios desse tipo forma a cascata e uma amostra é considerada "detectada" ao ser aprovada pelo último estágio, sendo que o n-ésimo estágio trata apenas as amostras que não foram rejeitadas nos n-1 estágios anteriores.



Figura 3.5: Cascata de classificadores.

Esse desenho em cascata do classificador permite que o esforço computacional empregado na avaliação de cada amostra seja proporcional à sua semelhança com o padrão buscado pois amostras obviamente negativas são descartadas nos primeiros estágios e amostras mais promissoras passam pela avaliação de mais estágios.

O desempenho de um classificador pode ser avaliado sobre um conjunto de amostras po-

sitivas e negativas. No contexto deste trabalho, amostras correspondem a regiões de imagens; se a região corresponde a uma instância do objeto que deve ser detectado, então trata-se de uma amostra positiva e, em caso contrário, trata-se de uma amostra negativa. Devido à diversidade de termos comumente utilizada para fazer referência ao desempenho de classificadores cabe explicitar os significados de alguns termos nesse texto:

$$HR = Recuperação = \frac{Amostras Positivas Classificadas Corretamente}{Total de Amostras Positivas}$$
(3.3)

$$FA = Taxa de Alarmes Falsos = \frac{Amostras Negativas Classificadas Erroneamente}{Total de Amostras Negativas}$$
(3.4)

Os valores mínimo/máximo admissíveis para HR e FA são parâmetros de treinamento e limitam a quantidade de falsos negativos e falsos positivos produzidos em cada estágio da cascata durante o treinamento. Como os falsos positivos de um estágio passarão para a avaliação dos próximos estágios da cascata e podem ser eliminados posteriormente, um valor alto de FA (entre 40% e 50%) é admissível. Já no caso dos falsos negativos, como os mesmos são descartados no momento em que são classificados como negativos, o HR da cascata piorará quanto maior for o número de estágios da mesma. Portanto valores altos de HR (entre 98% e 99,8%) devem ser exigidos a cada estágio para limitar o descarte equivocado de amostras positivas no classificador final.

Como em tempo de execução as amostras descartadas em um estágio não são mais processadas no estágio seguinte, o uso de uma cascata de classificadores se torna uma maneira natural de concentrar o processamento apenas em amostras promissoras, garantindo um melhor investimento da capacidade computacional.

#### 3.4 Boosting

Cada estágio da cascata de classificadores (seção 3.3) é constituído por um classificador ou uma pequena árvore de classificadores gerados através da técnica de *boosting* que possibilita a combinação de vários classificadores (tipicamente classificadores fracos) através de um esquema de votação para formar um classificador mais robusto. O *boosting* permite construir um classificador com níveis de recuperação (HR) e alarme falso (FA) arbitrariamente altos e baixos, respectivamente, bastando para isso acrescentar novos classificadores "fracos" em sua composição até que os índices desejados de HR e FA sejam atingidos no conjunto de treinamento.

Um classificador "fraco" apenas precisa ser melhor que o acaso (chance maior que 50% de acertar entre duas categorias) para que seja adequado ao uso em um algoritmo de *boosting*. No arcabouço para detecção rápida de objetos de Viola-Jones, um classificador fraco é formado por uma característica baseada em *haar-like features* e o *boosting* encarrega-se de combinar

o resultado de vários desses classificadores fracos em uma soma ponderada resultando em um classificador com maior poder de expressão.

O trabalho de Viola-Jones se baseia no algoritmo de *boosting* chamado *Discrete Adaboost* (Freund e Schapire (1997)) cuja idéia central será detalhada a seguir. Considere um conjunto  $C = \{(x_i, y_i) : i = 1, 2, ..., n\}$  de treinamento composto por amostras  $(x_i, y_i)$  onde  $x_i$  é uma amostra pertencente a um conjunto  $X \in y_i$  é uma classe pertencente ao conjunto  $\{-1, +1\}$ .

O Discrete Adaboost funciona em T rodadas e em cada uma delas um classificador fraco é selecionado para compor o classificador final. Para cada rodada t, vários classificadores fracos são treinados e, dentre eles, é selecionado um classificador representado pela função  $h_t : X \to \{-1, +1\}$  que minimiza o erro de classificação de uma quantidade de exemplos do conjunto de treinamento ponderando os resultados segundo uma distribuição de pesos  $P_t$ . Para t = 1, tem-se que  $P_1((x_i, y_i)) = 1/n$  para i = 1, 2, ..., n, igualando todos os pesos iniciais das amostras.

Após o treinamento do classificador  $h_t$ , seu erro associado é calculado como:

$$erro(h_t) = \sum_{i=1}^{n} P_t((x_i, y_i)) \mathbb{1}_{[h_t(x_i) \neq y_i]}.$$
 (3.5)

Em seguida um fator de ajuste  $\alpha_t$  é calculado em cada rodada e usado para atualizar a distribuição de pesos das amostras de treinamento, diminuindo a importância de amostras classificadas corretamente e aumentando a importância dos erros para a próxima rodada da seguinte forma:

$$\alpha_t = \ln\left(\frac{1 - erro(h_t)}{erro(h_t)}\right) \tag{3.6}$$

$$P_{t+1}((x_i, y_i)) = \frac{P_t((x_i, y_i))e^{\alpha_t \mathbf{1}_{[h_t(x_i) \neq y_i]}}}{Z_t}; i = 1, 2, \dots, n$$
(3.7)

onde  $Z_t$  é o fator de normalização calculado como

$$Z_t = \sum_{i=1}^n P_t((x_i, y_i)) e^{\alpha_t \mathbf{1}_{[h_t(x_i) \neq y_i]}}.$$
(3.8)

Essa normalização garante que  $P_{t+1}$  seja uma distribuição de pesos com soma unitária, ou seja,

$$\sum_{i=1}^{n} P_{t+1}((x_i, y_i)) = 1.$$
(3.9)

Ao final, o classificador forte H(x) é obtido como uma combinação linear dos classificadores  $h_t$  selecionados em cada rodada do algoritmo e ponderados por seus respectivos  $\alpha_t$  e submetidos a um limiar de classificação, resultando em:

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right).$$
(3.10)

Viola e Jones adaptaram o Adaboost de maneira a especificar como critério de parada um objetivo de erro e de alarmes falsos ao invés do número de rodadas (T). Além de construir o classificador forte que representará um estágio da cascata de processamento, o Adaboost termina por realizar a seleção das melhores *haar-like features* dentre o enorme número de possibilidades.

### Capítulo 4

# Abordagem Proposta para Deteção de códigos QR

A solução proposta nesse trabalho aborda o problema de detecção de códigos QR em cenas arbitrárias através da detecção de objetos baseada em componentes. Segundo esse tipo de abordagem, em um primeiro momento são buscadas em separado partes do objeto de interesse e posteriormente são agregados os resultados para avaliar se são partes compatíveis entre si do padrão buscado. A motivação e os detalhes de formulação são apresentados nesse capítulo.

#### 4.1 Visão Geral da Abordagem Proposta

A detecção de códigos QR como um todo é tarefa complicada pois a organização estrutural interna desses códigos de barras pode variar de acordo com a quantidade de dados codificada e a versão selecionada para comportá-la. A Figura 2.4 apresentada no capítulo 2 ilustra essa variação de estrutura interna ao mostrar, por exemplo, como a proporção entre a área ocupada pelos FIPs e a área total do código QR se altera de versão para versão. Nesse sentido os códigos QR são considerados objetos "com formato não rígido".

Além disso a extração de características da imagem na região do código como, por exemplo, histogramas ou caracterização no domínio de frequência, podem ser fortemente influenciadas pela resolução da imagem, escala do código e ruído.

Por outro lado, os FIPs são tipicamente as maiores estruturas dentro de um código, têm forma rígida e aparecem exatamente nos três cantos de todos os códigos QR. Essas características sugerem que uma abordagem de detecção de códigos QR baseada em componentes (FIPs nesse caso) tem boa perspectiva de sucesso. A ideia principal da detecção baseada em componentes, explorada em Mohan *et al.* (2001), é realizar a detecção de componentes rígidos de um objeto complexo, eventualmente de formato não rígido, seguida da análise da posição relativa dos supostos componentes detectados. Em alguns casos o arranjo dos componentes detectados pode ser submetido a restrições geométricas determinadas analiticamente, como é o caso dos FIPs em códigos QR. Nos casos em que a posição relativa ideal das partes detectadas não segue uma regra pré-estabelecida a aprendizagem de máquina tem sido usada a exemplo de Agarwal *et al.* (2004) e também Felzenszwalb *et al.* (2010) que usa um modelo baseado em partes deformáveis.

O propósito dos FIPs é propriamente ajudar na determinação da localização e do contorno dos códigos QR e, portanto, os mesmos são usados dessa forma pelos algoritmos de decodificação existentes no mercado. Frequentemente tais algoritmos realizam, em uma primeira etapa, a determinação da posição do símbolo com base em linhas de escaneamento distribuídas de maneira regular ao longo da imagem em busca de padrões correspondentes a FIPs (veja a Figura 2.3). Contudo, ao tratar de imagens adquiridas ao acaso, a hipótese de que o código ocupa posição de destaque na imagem não é válida e a busca por FIPs ao longo de linhas de escaneamento se torna inviável pois as mesmas teriam de ser muito próximas para possibilitar a detecção de FIPs relativamente pequenos. Tal proximidade das linhas de escaneamento e a grande variação dos possíveis tamanhos de FIPs a serem buscados tem o potencial de aumentar o custo computacional desse tipo de abordagem e dificulta a obtenção de resultados precisos.

A motivação de técnicas baseadas em componentes é detectar com eficácia objetos que apresentam componentes de pouca variação local mas que variam consideravelmente quando considerados em conjunto em diferentes instâncias do mesmo tipo de objeto. Sendo assim, a abordagem proposta está dividida em dois estágios descritos a seguir:



Para o primeiro estágio, é proposta a utilização da técnica desenvolvida por Viola-Jones para detecção rápida de objetos, descrita no Capítulo 3.

O segundo estágio consiste de um algoritmo que impõe uma série de restrições geométricas ao arranjo de regiões detectadas como resultado do primeiro estágio. Ao final, uma combinação de 3 detecções com tamanhos similares arranjadas de maneira a formar os vértices de um símbolo código QR é considerada um resultado positivo.

#### 4.2 Detecção de FIPs usando Viola-Jones

A busca de códigos QR em imagens com enquadramento arbitrário elimina a possibilidade de estabelecer qualquer tipo de restrição relativa ao tamanho desses símbolos na cena a ser processada. Essa característica aliada ao requisito de processamento em tempo real torna necessária a busca de uma abordagem para detecção de FIPs que seja rápida e invariante à escala. Os FIPs, isoladamente, são objetos de formato "rígido" e portanto são totalmente adequados à detecção através do arcabouço para detecção rápida de objetos desenvolvido por Viola e Jones.

Outro fator importante a ser considerado é que, por tratar-se a abordagem proposta de uma deteção em dois estágios, é importante que exista a possibilidade de influir no desenho do primeiro estágio para ajustar a relação entre recuperação e falsos positivos da maneira mais vantajosa ao resultado como um todo e não apenas aos resultados do primeiro estágio.

No primeiro estágio é desejável que sejam maximizados os verdadeiros positivos (TP) enquanto a importância do número de falsos positivos (FP), embora exista, é secundária. Dessa forma é importante a existência de algum mecanismo durante o projeto do primeiro estágio para controlar TP e FP para que o processo de detecção atenda aos requisitos mencionados.

O arcabouço para detecção rápida de objetos de Viola e Jones é invariante à escala, é muito rápido e tem grande flexibilidade nos parâmetros de treinamento para possibilitar os ajustes necessários de desempenho.

### 4.3 Agregação de Candidatos a FIP para Detecção de Códigos QR

Uma descrição formal do problema de agregação de candidatos a FIP é apresentada e, a partir dela, um algoritmo é proposto e descrito em detalhes.

#### 4.3.1 Formulação do Problema

Cada candidato a FIP detectado no primeiro estágio é representado por uma tupla (T, x, y) onde T é o lado da região quadrada detectada e x e y representam as coordenadas do centro desse quadrado na imagem.

Em um cenário ideal livre de ruído ou distorção perspectiva, quaisquer dois FIPs de um mesmo código QR têm o mesmo tamanho. Além disso, a estrutura dos códigos QR limita a distância possível entre dois FIPs de um mesmo código QR a valores mínimos e máximos em número de módulos que, por sua vez, têm relação fixa com o tamanho de um FIP. Um FIP corresponde a  $7 \times 7$  módulos (ou a  $9 \times 9$  módulos ao considerar-se a região branca em torno dele), como pode ser observado na Figura 4.1. A menor versão de código QR (versão 1) pode acomodar 21 módulos em largura (ver Fig. 4.1) enquanto a maior versão (versão 40) apresenta largura de 177 módulos. Dessa maneira, o número mínimo de módulos entre

os centros de dois FIPs adjacentes é 21 - 7 = 14, e o máximo é 177 - 7 = 170. Portanto, a distância de dois FIPs de lado T adjacentes em um código QR é limitada inferiormente por  $(T/9)14 \approx 1.6T$  e superiormente por  $(T/9)170 \approx 19T$ . Uma vez que FIPs podem encontrar-se em cantos opostos do código pela diagonal, a distância máxima possível para dois FIPs que façam parte do mesmo símbolo é  $19\sqrt{2}T$ .



Figura 4.1: Um código QR de versão 1 é composto por  $21 \times 21$  módulos.

Além disso os segmentos de reta que unem o centro dos FIPs em um código QR são dois a dois ortogonais entre si (dois lados da região quadrada) ou formam um ângulo de 45° (um dos lados e a diagonal da região quadrada). Dados dois FIPs em um código QR, existem seis possíveis posições para o terceiro FIP como ilustrado na Figura 4.2.



**Figura 4.2:** Dados dois FIPs (indicados como 1 e 2) de um mesmo código QR, as seis possíveis posições que um terceiro FIP pode ocupar são representadas pelos círculos negros.

Com base nessas observações foram propostos critérios para a avaliação do conjunto de candidatos a FIPs.

**Definição 1** (Critério de Tamanho). Dois candidatos a FIP  $(T_1, x_1, y_1)$  e  $(T_2, x_2, y_2)$  satisfazem o critério de tamanho se e somente se  $T_1 = T_2$ .

**Definição 2** (Critério de Distância). Dois candidatos a FIP  $(T, x_1, y_1)$  e  $(T, x_2, y_2)$  do mesmo tamanho T satisfazem o critério de distância se e somente se

$$1.6 T \le dist((x_1, y_1), (x_2, y_2)) \le 19\sqrt{2} T$$

Sejam  $F_1 = (T_1, x_1, y_1)$  e  $F_2 = (T_2, x_2, y_2)$  dois candidados a FIP sendo que  $\overline{F_1 F_2}$  denota o segmento de reta que une  $(x_1, y_1)$  a  $(x_2, y_2)$ , e  $|\overline{F_1 F_2}|$  representa seu comprimento. Note que  $\overline{F_2 F_1} = \overline{F_1 F_2}$ .
**Definição 3** (Critério de Geometria). Sejam  $\overline{F_1 F_2}$  e  $\overline{F_1 F_3}$  dois segmentos de reta com extremidade comum  $(x_1, y_1)$ ,  $\theta$  o ângulo entre eles, e  $P = \frac{\min\{|\overline{F_1 F_2}|, |\overline{F_1 F_3}\}}{\max\{|\overline{F_1 F_2}|, |\overline{F_1 F_3}|\}}$ . Então,  $\{\overline{F_1 F_2}, \overline{F_1 F_3}\}$  satisfaz o critério de geometria se e somente se  $(i) \ \theta = 90^\circ \ e \ P = 1 \ (i.e., |\overline{F_1 F_2}| = |\overline{F_1 F_3}|), \ ou$  $(ii) \ \theta = 45^\circ \ e \ P = \frac{\sqrt{2}}{2}.$ 

Se um conjunto de três candidatos a FIPs realmente pertence a um mesmo código QR então eles devem satisfazer, dois a dois, os critérios de tamanho e distância e os três segmentos de retas com extremidades em seus centros devem satisfazer, dois a dois, o critério de geometria. Portando, o problema de encontrar candidatos a códigos QR na imagem pode ser entendido como uma questão de encontrar subconjuntos de três FIPs que satisfaçam as três condições apresentadas acima.

A formulação proposta para a solução desse problema é baseada em grafos. Os candidatos a FIP são representados como vértices e dois vértices são ligados por uma aresta se seu respectivo par de FIPs satisfaz os critérios de tamanho e distância. Deve-se notar porém que não são todos os trios de candidados a FIP que satisfaçam, dois a dois, os critérios de tamanho e distância que podem ser identificados como membros de um mesmo código QR. Portanto, a solução do problema não corresponde a todos os cliques de tamanho três no gráfico mas sim é necessário verificar adicionalmente se os segmentos de retas correspondentes às arestas nesse conjunto satisfazem o critério de geometria.

**Proposição 1.** Seja  $\{F_1, F_2, F_3\}$  um conjunto de candidatos a FIP que satisfaçam, dois a dois, os critérios de tamanho e distância. Então,  $\{\overline{F_1 F_2}, \overline{F_1 F_3}\}$  satisfaz o critério de geometria se e somente se  $\{\overline{F_2 F_1}, \overline{F_2 F_3}\}$  e  $\{\overline{F_3 F_1}, \overline{F_3 F_2}\}$  também satisfazem o critério de geometria.

Demonstração. : Se os segmentos  $\{\overline{F_1 F_2}, \overline{F_1 F_3}\}$  satisfazem o critério de geometria então eles representam os dois lados de um quadrado ou um lado do quadrado em sua diagonal. No caso de representarem dois lados de um quadrado o terceiro segmento de reta  $\overline{F_2 F_3}$  é a diagonal e obviamente satisfaz a condição (ii) do critério de geometria em relação a  $\overline{F_1 F_2}$  e também a  $\overline{F_1 F_3}$ . No outro caso, em que um deles representa um lado do quadrado e o outro sua diagonal, obviamente o terceiro segmento de reta é um lado do quadrado que satisfaz a condição (i) do critério de geometria com o segmento que representa um lado do quadrado enquanto satisfaz a condição (ii) do critério de geometria com o segmento que representa a diagonal do quadrado.

Uma consequência importante dessa proposição é que, dados três candidatos a FIP que satisfaçam, dois a dois, os critérios de tamanho e distância basta que apenas um dos pares de segmentos de retas associados a eles satisfaça o critério de geometria para determinar que tais candidatos em conjunto correspondem a um candidato a código QR.

Além disso, em uma condição ideal, dado que um FIP é parte de um código QR, ele não pode ser parte de outro. Dessa forma, tão logo um trio de candidatos que satisfizesse essa condição fosse identificado, tais candidatos poderiam ser removidos do grafo junto com as arestas a eles adjacentes. Contudo, podem existir situações como a ilustrada na Figura 4.3 na qual o primeiro trio de candidatos a FIP detectado é um falso positivo. Para evitar o descarte dos FIPs equivocadamente, que anularia a possibilidade de detecção do código verdadeiro, uma aresta só pode ser removida após a análise de todas as que lhe forem adjacentes.



**Figura 4.3:** Uma aresta que pode ser parte de dois subconjuntos de três candidatos a FIP satisfazendo os critérios.

Na prática as medidas de tamanho e distância obtidas a partir de uma imagem não são precisas em decorrência da qualidade do processo de aquisição, da resolução da imagem, rotação e distorções perspectivas. Faz-se necessária a elaboração de critérios que possibilitem alguma tolerância para acomodar imprecisões decorrentes dessas condições. Sejam dois candidatos a FIP  $(T_1, (x_1, y_1))$  e  $(T_2, (x_2, y_2))$ . Tolerâncias podem ser acomodadas reformulando os critérios conforme descritos a seguir:

- (tamanho) Considere  $T = \max\{T_1, T_2\}$  e  $t = \min\{T_1, T_2\}$ . É permitida uma variação de tamanho entre dois candidatos a FIP até um percentual, ou seja,  $T t \leq \varepsilon T$ . Por exemplo  $\varepsilon = 0.2$  indica que a tolerância para a diferença de tamanho entre dois candidatos a FIP é 20%.
- (distância) Versões grandes de códigos QR são raramente encontradas e portanto a distância máxima pode ser estabelecida de acordo com a maior versão de código que se pretende encontrar diminuindo assim a chance de encontrar falsos positivos. Nesse trabalho não foi limitada a versão a ser detectada, ou seja, a maior versão possível é a 40. Dessa maneira a distância entre dois candidatos a FIP deve respeitar os limites 1.6 T ≤ dist((x<sub>1</sub>, y<sub>1</sub>), (x<sub>2</sub>, y<sub>2</sub>)) ≤ 19√2 T, onde T = max{T<sub>1</sub>, T<sub>2</sub>}.
- (geometria): uma pequena variação no ângulo e no comprimento dos segmentos de reta que unem os centros dos candidatos a FIP deve ser admitida. Essa tolerância pode ser parametrizada através de um parâmetro único d e da adaptação das condições do critério de geometria da seguinte forma: (i)  $|90^{\circ}-\theta| \leq d \in 1 \frac{\cos(45^{\circ}-d) \cos(45^{\circ}+d)}{\cos(45^{\circ})} \leq P \leq 1$  ou (ii)  $|45^{\circ}-\theta| \leq d \in \cos(45^{\circ}+d) \leq P \leq \cos(45^{\circ}-d)$ .

## 4.3.2 Algoritmo de Agregação de FIPs

A solução proposta para agregação de FIPs é apresentada no Algoritmo 1. As linhas 1 a 3 correspondem à criação do conjunto de vértices. As linhas 4 a 13 correspondem à criação

de arestas toda vez que um par de vértices satisfaz os critérios de tamanho e distância com suas respectivas tolerâncias. A estrutura de laço da linha 14 à linha 23 é a parte principal do algoritmo onde os cliques de tamanho três cujas arestas satisfazem, duas a duas, o critério de geometria são encontrados. Para cada aresta (u, v), todas as arestas (u, v') que são adjacentes à sua extremidade u são verificadas. Para cada par de arestas desse tipo (i.e.,  $\{(u, v), (u, v')\}\}$ ), a existência da terceira aresta (v, v') é verificada. Se a terceira aresta não está presente, conclui-se que os dois vértices v and v' não satisfazem o critério de tamanho ou distância ou que a terceira aresta já foi processada (nesse caso, a tupla  $\{u, v, v'\}$  já foi encontrada e consta na lista de saída). Se a terceira aresta (v, v') está presente e se as duas primeiras satisfazem o critério de geometria então o subconjunto de três vértices  $\{u, v, v'\}$  é adicionado à lista de saída. Note-se que não é necessário examinar arestas adjacentes à outra extremidade (v)dado que essas serão avaliadas posteriormente. O algoritmo termina quando todas as arestas forem processadas. Dessa forma, pode-se concluir que a complexidade do Algoritmo 1 para uma entrada com n candidados a FIP é  $O(n^3)$ .

-	Algoritmo 1: Algoritmo de agregação de candidatos a FIP.									
-	<b>Entrada</b> : uma lista de candidatos a FIP, cada um representado por uma tupla $\overline{D}_{1}$									
	$F_i = (T_i, x_i, y_i)$ , e parâmetros de tolerância $\varepsilon$ e d									
	<b>Saída</b> : uma lista de subconjuntos da forma $\{F_i, F_j, F_k\}$ satisfazendo os critérios de tamanho, distância e geometria em suas respectivas tolerâncias.									
1	para cada $F_i$ faça									
<b>2</b>	crie um vértice $v_i$ e marque-o como "não processado"									
3	fim									
4	enquanto existem vértices "não processados" u restantes faça									
5	para cada vértice "não processado" $v, v \neq u$ , faça									
6	se $\{u, v\}$ satisfaz os critérios de tamanho e com tolerância $\varepsilon$ então									
7	$\mathbf{se} \{u, v\}$ satisfaz o critério de distância então									
8	crie a aresta $(u, v);$									
9	fim fim									
10	fim									
11	fim									
<b>12</b>	marque o vértice $u$ como "processado";									
13	fim									
14	enquanto existem arestas $(u, v)$ restantes faça									
15	para cada aresta $(u, v'), v' \neq v$ , faça									
16	se existe aresta $(v, v')$ então									
17	se arestas $(u, v)$ e $(u, v')$ satisfazem o critério de geometria com tolerância									
	d entao									
18	adicione $\{u, v, v'\}$ à lista de saída									
19										
20										
21	fim									
22	Remova a aresta $(u, v)$ ;									
23	fim									

# Capítulo 5

# Treinamento e avaliação de detector de FIPs

O treinamento de um detector baseado na técnica de Viola-Jones demanda, além de grandes conjuntos de exemplos, a determinação de parâmetros de desenho da cascata que influenciarão a qualidade dos resultados alcançados. A determinação dos parâmetros de treinamento mais adequados à detecção de FIPs bem como o estabelecimento de um conjunto de treinamento apropriado são essenciais para o bom emprego da técnica nesse novo domínio de aplicação.

São apresentados nesse capítulo detalhes das ferramentas utilizadas, do treinamento de classificadores através de *boosting* e da metodologia seguida para a determinação de parâmetros de treinamento adequados a serem usados na criação de um detector de FIPs.

## 5.1 Implementação utilizada

Esta seção descreve detalhes das ferramentas utilizadas para o treinamento de classificadores, para a geração automatizada de amostras de treinamento artificiais através de variações de imagens originais e para a medição de desempenho dos classificadores.

## 5.1.1 Utilitários do OpenCV

As ferramentas utilizadas durante o treinamento e avaliação dos classificadores fazem parte da biblioteca Opencv (Bradski 2000). São elas:

**Treinamento de classificador:** O utilitário opencv-haartraining é uma implementação de treinamento de um classificador constituído por uma cascata de classificadores segundo a proposta de Viola-Jones. Foi incorporado à biblioteca após o trabalho realizado por Lienhart e Maydt (2002) e admite uma grande quantidade de variações através de seus parâmetros de ajuste. Uma descrição detalhada desses parâmetros é feita na seção 5.1.3. Criação de amostras de treinamento: Como o número de amostras de treinamento necessárias para o treinamento de uma cascata de classificadores é elevado, o utilitário opencv-createsamples fornece uma maneira de ampliar de modo artificial a quantidade de amostras disponíveis. Por meio de sua aplicação é possível criar diferentes versões de uma mesma imagem através da aplicação de variações aleatórias e limitadas de ângulo (em  $x, y \in z$ ) e iluminação.

Avaliação automatizada de desempenho: Medir e comparar o desempenho de um classificador em imagens é uma tarefa demorada e propícia a erros quando realizada de maneira manual. O utilitário opencv-performance provê uma forma de avaliar o desempenho de um classificador construído a partir de opencv-haartraining, fornecendo como resultado o número de acertos, erros e falsos positivos encontrados no conjunto de imagens bem como uma curva ROC que reflete o efeito de se variar o número mínimo de detecções sobrepostas exigidas durante a detecção (ND, detalhes na seção 5.1.4).

### 5.1.2 Algoritmo de *Boosting*

O algoritmo de *boosting* usado na abordagem proposta é ligeiramente distinto do utilizado por Viola e Jones e é denominado *Gentle Adaboost*. Os experimentos de Lienhart e Maydt (2002) mostraram que esse algoritmo obteve desempenho superior ao *Discrete Adaboost* utilizado por Viola e Jones.

A principal diferença entre *Discrete Adaboost* e *Gentle Adaboost* é que no segundo os classificadores devem produzir valores contínuos de saída e, portanto, a atualização dos pesos é feita de maneira diferente. No *Gentle Adaboost* os pesos são atualizados da seguinte forma:

$$h_t(x) \in [0, 1] \text{ (contínuo)} \tag{5.1}$$

$$P_{t+1}((x_i, y_i)) = \frac{P_t((x_i, y_i))e^{-y_i h_t(x_i)}}{Z_t}; i = 1, 2, \dots, n.$$
(5.2)

onde  $Z_t$  é o fator de normalização calculado como

$$Z_t = \sum_{i=1}^n P_t((x_i, y_i)) e^{-y_i h_t(x_i)}.$$
(5.3)

A seção 3.4 contém uma descrição completa da notação utilizada.

## 5.1.3 Parâmetros de Treinamento

Vários parâmetros de treinamento podem ser configurados para alterar a forma do classificador resultante de acordo com requisitos de domínio ou de desempenho. Os parâmetros a seguir serão considerados em nossa análise:

- Conjunto de Características (MODE) : As duas possibilidades são o conjunto básico (Figura 3.1) e o conjunto ampliado (Figura 3.2), que contempla rotações a 45° dos protótipos básicos.
- Simetria (SYM): Quando a forma de interesse é simétrica, o algoritmo de treinamento pode ser instruído a considerar apenas metade das imagens de entrada para reduzir o tempo de treinamento.
- Topologia do Classificador (MTS): É possível permitir que existam ramificações na cascata, convertendo-a de uma árvore CART degenerada para uma árvore CART simples. Esse parâmetro estabelece um limite máximo para as subdivisões dessa árvore CART.
- Número de divisões do classificador fraco (NS): Limita o número de subdivisões presentes em um classificador fraco. Um classificador fraco em sua forma mais simples consiste em uma única haar-like feature associada a um limiar binário (NS=1) que será combinado com outros classificadores similares em um esquema de votação para constituir um estágio. É possível permitir a tais classificadores fracos a representação de relacionamentos mais sofisticados ao permitir que eles tomem uma forma mais sofisticada como uma árvore CART com um pequeno e limitado (NS>1) número de ramificações.
- Mínima taxa de acerto e máxima taxa de alarme falso (HR/FA): Cada estágio na cascata é treinado para respeitar um limite inferior para taxa de acerto e superior para alarme falso. Durante o treinamento, classificadores fracos são adicionados à composição do estágio até que o mesmo se enquadre simultaneamente nos dois limites mencionados. Um valor demasiadamente baixo de tolerância para alarmes falsos pode ocasionar a geração de estágios excessivamente complexos, eliminando os benefícios do processamento em cascata. Uma exigência muito alta para a taxa mínima de acerto pode ter efeitos semelhantes.
- Número de amostras de treinamento (SAMPLES): Quantidade de amostras positivas utilizadas no treinamento.
- Tamanho das amostras de treinamento (SIZE): Durante o treinamento e a aplicação do classificador, as amostras são redimensionadas a um tamanho fixo. Para a detecção de faces, Viola e Jones consideraram 20 × 20 um tamanho adequado.

A determinação dos melhores parâmetros a serem usados no treinamento da cascata para detecção de FIPs, se fosse feita de maneira exaustiva, implicaria na investigação de um número muito grande de possibilidades uma vez que cada um dos 7 parâmetros considerados pode assumir diversos valores e as possibilidades de combinação são muitas. Dessa forma, a determinação foi feita com base na observação dos efeitos da variação de um único parâmetro nos resultados, enquanto todos os outros foram mantidos constantes conforme descrito em Belussi e Hirata (2011). Esse mesmo tipo de abordagem foi usado por Lienhart *et al.* (2002) ao investigar parâmetros adequados ao domínio de detecção de faces.

A seção 5.2.2 detalha o procedimento usado para determinação dos parâmetros adequados à detecção de FIPs.

#### 5.1.4 Avaliação do Detector de FIP

Um classificador treinado usando a implementação de opencv-haartraining do OpenCV 2.0 pode ter seu desempenho facilmente avaliado através do utilitário opencvperformance. Essa avaliação depende da disponibilização de imagens de teste com gabaritos associados na forma de arquivos que descrevem a posição e tamanho dos objetos de interesse. Tais imagens, ao invés de obtidas e marcadas manualmente, podem ser geradas através do utilitário opencv-createsamples que opera inserindo imagens contendo apenas o objeto de interesse em imagens de fundo (que não devem conter nenhuma instância de tal objeto) e gerando automaticamente o arquivo de gabarito que descreve a posição e tamanho da imagem inserida. Essa ferramenta aliada ao uso de opencv-performance mostra-se de grande utilidade na automação de testes de desempenho. A Figura 5.1 mostra um exemplo de imagem construída pelo opencv-createsamples segundo o procedimento descrito.



Figura 5.1: Uma imagem de fundo com a inserção de um FIP.

O desempenho de um detector em um conjunto de imagens de teste pode ser medido por

uma combinação do número de verdadeiros positivos (TP), falsos positivos (FP), verdadeiros negativos (TN) e falsos negativos (FN). Além disso, as medidas de recuperação e precisão são usadas para expressar o desempenho do classificador. A recuperação é defininda como a fração de amostras positivas que são efetivamente detectadas como positivas, ou seja,

$$\operatorname{Recuperação} = \frac{TP}{TP + FN} \,. \tag{5.4}$$

A precisão é definida como a fração das amostras detectadas como positivas pelo classificador que realmente correspondem a amostras positivas, sendo então calculada como

$$\operatorname{Precisão} = \frac{TP}{TP + FP} \,. \tag{5.5}$$

Dois importantes parâmetros a serem considerados no momento da aplicação do detector são o fator de escalonamento (SF) e o número mínimo de detecções sobrepostas (ND) uma vez que seus ajustes podem influenciar significativamente a qualidade dos resultados e o tempo de processamento.

Fator de Escalonamento (SF): Durante o processo de detecção, uma janela deslizante de processamento é utilizada percorrendo a imagem pixel a pixel. Para cada posição possível na imagem, várias amostras são obtidas aumentando as dimensões da janela deslizante de processamento a partir de um valor definido na chamada do detector até o maior valor possível de modo que a janela não exceda a área da imagem a partir daquela posição. Toda amostra capturada pela janela de processamento é redimensionada para o tamanho de janela utilizada durante o treinamento do detector de forma que as dimensões dessa janela de treinamento constituem um limitante inferior para o parâmetro usado na chamada desse detector. Isto implica que um dado detector jamais encontrará objeto de tamanho menor que o usado durante seu treinamento. O fator de escalonamento determina a taxa a que as dimensões da janela deslizante de processamento será aumentada, por exemplo: Um fator de escalonamento SF=1.10 significa que as dimensões da janela de processamento será aumentadas de 10% a cada rodada de aquisição de amostras em cada posição. O fato de que todas as amostras adquiridas pela janela deslizante de processamento são redimensionadas para um tamanho fixo garante a invariância à escala desse processo de detecção.

Número mínimo de detecções sobrepostas (ND): Frequentemente várias amostras que correspondem a pequenas variações de posição em torno do objeto a ser detectado na imagem são apontadas como positivas, ou seja, são consideradas detecções distintas. Esses casos são muito frequentes e tornam conveniente o estabelecimento de um critério para a consolidação dessas detecções em uma única detecção que de fato as represente como conjunto. Naturalmente, quanto maior o número de detecções sobrepostas em determinada vizinhança, maior a evidência de que naquela posição da imagem encontra-se, de fato, o objeto de interesse. Portanto o número mínimo de detecções sobrepostas (ND) é estabelecido como critério para determinar se um dado conjunto de detecções próximas deve ser considerado como uma detecção verdadeira, ou seja, apenas conjuntos de detecções sobrepostas de cardinalidade maior que ND serão considerados uma detecção.

## 5.2 Determinação dos parâmetros de treinamento

A determinação dos parâmetros de treinamento do detector foi realizada através da variação individual de todos eles e a verificação de sua influência no desempenho do classificador resultante. Os parâmetros são descritos na seção 5.1.3 e o resultado da comparação e escolha dos parâmetros é apresentado na seção 5.2.2.

Os treinamentos e testes dos detectores de FIPs foram realizados com o auxílio das ferramentas opencv-haartraining e opencv-performance.

### 5.2.1 Conjunto de Dados

O conjunto total de amostras utilizadas durante a determinação de parâmetros de treinamento e detecção é composto de N = 380 exemplos de FIPs recortados de diversas imagens obtidas em diferentes condições de local, enquadramento e iluminação. O conjunto de amostras negativas é composto de 1500 imagens que não contêm códigos QR.

As amostras de treinamento devem ser adquiridas com cautela de maneira a preservar a posição relativa dos padrões na janela de recorte pois as *haar-like features* são muito sensíveis ao posicionamento das mesmas relativamente à janela de processamento. Há também que se observar o aspecto das amostras que deve ser idêntico ao especificado para treinamento (por ex.  $16 \times 16$ ) para que não hajam distorções no processo.

Para compor o conjunto de treinamento foram usadas 285 das 380 amostras de FIPs. Essas amostras foram submetidas a variações aleatórias e limitadas de luminosidade e de ângulo nos eixos x, y (perspectiva) e z (planar) através da ferramenta opencv-createsamples, ampliando o conjunto de treinamento e possibilitando aferir o efeito desse aumento nos resultados obtidos. As variações foram limitadas a ±80 em intensidade, ±23° em  $x e y e \pm 34°$ em z. As amostras negativas foram divididas igualmente entre treinamento e teste sendo 750 para cada finalidade.

O conjunto de teste foi construído a partir das 95 amostras de FIPs restantes, submetidas a variação aleatória limitada a  $\pm 80$  em intensidade,  $\pm 23^{\circ}$  em x,y e em z. A partir dessas 95 amostras foram geradas 750 variações que foram posicionadas em cada uma das imagens negativas de teste, formando assim um conjunto artificial de imagens de teste com 750 exemplares.

A geração automatizada das imagens de teste, feita através do utilitário opencv-createsamples, tem a vantagem de produzir exemplares marcados, ou seja, imediatamente aptos a serem processados pelo utilitário opencv-performance que automatiza o processo de testes.

## 5.2.2 Determinação experimental de parâmetros de treinamento

O treinamento de um detector de FIPs constituído por uma cascata de classificadores conforme proposto por Viola-Jones é um processo que admite uma série de variações e ajustes. A escolha adequada dos parâmetros responsáveis por tais ajustes, chamados de parâmetros de treinamento, é fator determinante de sucesso no emprego dessa técnica.

A metodologia para a determinação dos parâmetros é descrita em 5.1.3 e baseia-se na variação individual dos parâmetros de treinamento.

A Figura 5.2 mostra a ordem na qual os parâmetros foram considerados. No início, para avaliar o conjunto de características (MODE) a determinação dos valores iniciais dos outros parâmetros foi feita com base em observações empíricas da seguinte forma: SYM=N (Assimétrico), MTS=0 (topologia de cascata), NS=2, FA=0.5, SAMPLES=4000 (positivas e negativas, 4000 cada), e SIZE= $20 \times 20$ . Para os parâmetros restantes, os que já haviam sido avaliados foram fixados nos valores escolhidos enquanto os outros permaneceram com os valores iniciais (com exceção, obviamente, do parâmetro sob avaliação). Para todos os treinamentos HR foi mantido fixo em 0.998 e ST em 15, exceto quando FA estava sendo avaliado pois a variação do número de estágios afeta diretamente os valores de HR e FA esperados da cascata ao final do treinamento, conforme detalhado na seção 5.1.3. Após 7 rodadas de treinamento (correspondentes aos parâmetros estudados), o conjunto de parâmetros escolhido para a detecção de FIPs é composto pelos valores dos nós escuros na Figura 5.2.

É importante salientar que a taxa de recuperação (grafada *Recall* em algumas figuras) e FP (falsos positivos) nos resultados de cada fase da determinação de parâmetros são relativos ao conjunto de teste, totalmente independente do conjunto de treinamento.

Uma discussão adicional sobre a escolha dos parâmetros é feita na seção 6.6.

A Figura 5.3 apresenta os resultados dos 20 classificadores gerados durante a determinação dos valores de parâmetros de treinamento ilustrada na Figura 5.2. Alguns resultados estão destacados através de setas numeradas, a saber:

- Seta 1 Classificador com a melhor relação Recuperação/FP. O número de alarmes falsos reduzidos se deve à imposição de um limite mais baixo para alarmes falsos em cada estágio da cascata (FA = 0, 4).
- Seta 2 Classificador com o melhor valor de Recuperação. Corresponde a uma cascata com classificadores mais complexos, permitindo até 8 divisões na estrutura dos classificadores fracos que o compõem (NS = 8). Convém ressaltar que um classificador muito mais simples (NS = 1) apresentou menos de 5% de diferença no resultado de Recuperação.
- Seta 3 Classificador que utiliza o conjunto ampliado de protótipos de características. Seu baixo desempenho pode indicar que ao aumentar substancialmente o espaço de classificadores possíveis o treinamento convergiu rapidamente mas não resultou em um classificador suficientemente geral, prejudicando o seu desempenho final.



**Figura 5.2:** Escolha de parâmetros: em cada nível o nó de cor escura corresponde ao valor escolhido.

**Seta 4** Classificador treinado com amostras de tamanho 6 × 6. Seu desempenho ruim indica que as amostras de treinamento com tamanho tão reduzido não contêm informação suficiente para uma classificação efetiva. Essa condição pode ter sido agravada pelo fato de que a existência de simetria estava sendo presumida (MODE=SYM).

Ao final dos experimentos o conjunto de parâmetros de treinamento escolhido é o que corresponde ao classificador "i". Apesar de apresentar FP comparativamente alto, os parâmetros utilizados para treinamento do classificador "i" foram escolhidos pois possibilitam a detecção de FIPs consideravelmente menores  $(16 \times 16)$  que a maior parte dos classificadores treinados (utilizando amostras de  $20 \times 20$ ). Em situações práticas, onde possivelmente existam FIPs com dimensões reduzidas, a escolha do menor tamanho de amostra de treinamento possível implica em um aumento da recuperação esperada acompanhado de certo aumento



**Figura 5.3:** Recuperação  $\times$  FP para classificadores obtidos com diferentes parâmetros de treinamento.

de FP que deverá ser contido pelo segundo estágio. Portanto os parâmetros de treinamento eleitos foram:

- Conjunto de Características (MODE) : Básico;
- Simetria (SYM): Simétrico;
- Topologia do Classificador (MTS): Cascata;

- Número de divisões do classificador fraco (NS): 1;
- Máxima taxa de alarme falso (FA): 0.50;
- Número de amostras de treinamento: 4000;
- Tamanho das amostras de treinamento:  $16 \times 16$ .

Pode-se chamar de características principais as *haar-like features* escolhidas durante os primeiros estágios da cascata de classificadores uma vez que representam o padrão que melhor distingue as amostras positivas das negativas logo no início do treinamento. É interessante notar que, quando a simetria do objeto a detectar é presumida durante o treinamento, os classificadores resultantes apresentam características principais diferentes se comparados ao classificadores gerados sem tal presunção. A Figura 5.4 apresenta as duas características principais para os dois casos.



**Figura 5.4:** Duas características principais para MODE=SYM (linha superior) e MODE=NONSYM (linha inferior).

O número médio de características usadas nos classificadores treinados foi de 110 e, independentemente dos parâmetros usados no treinamento, em geral não houve variação significativa desse número. As exceções foram os classificadores treinados com amostras de tamanho  $8 \times 8$  e  $6 \times 6$  que demandaram um número muito maior de características para atingir os níveis exigidos de acertos e alarmes falsos.

## 5.2.3 Aumento das Amostras de Treinamento

O efeito do aumento artificial das amostras no desempenho do detector resultante foi avaliado detalhadamente através do treinamento de 10 detectores distintos que têm em comum todos os parâmetros de treinamento exceto o número de amostras positivas utilizadas. O fator de multiplicação das amostras originais variou de 1 a 28, resultando em conjuntos de amostras positivas com tamanhos que variam entre 285 e 7980.

A Figura 5.5 ilustra uma imagem e suas correspondentes variações obtidas através desse processo.



**Figura 5.5:** Exemplos de amostras geradas a partir de variações aleatórias de uma amostra positiva.

O processo de treinamento não é determinístico o que pode ser constatado por diferenças de desempenho que existem entre dois classificadores que tenham sido treinados com os mesmos parâmetros e os mesmos conjuntos de amostras positivas e negativas para treinamento. Isso ocorre em virtude de que as amostras negativas usadas no treinamento são, de fato, constituídas por recortes aleatórios das imagens que foram fornecidas no conjunto de amostras negativas de treinamento. De forma a minimizar a influência de eventuais variações de desempenho e isolar a influência do aumento artificial do número de amostras positivas, cada um dos 10 detectores foi treinado 5 vezes usando os mesmos conjuntos de dados. Foi então observada a média do desempenho das 5 instâncias de cada classificador.

A Figura 5.6 mostra o desempenho de classificadores obtidos usando um número fixo de amostras negativas (4000) e variados tamanhos do conjunto de amostras positivas. Todas as amostras positivas foram obtidas através de variações aleatórias de um conjunto base de 285 FIPs. Os parâmetros de treinamento foram definidos de acordo com os parâmetros do classificador k na Figura 5.3 com exceção do número de estágios que foi elevado para 17 (ST=17). Os valores de Recuperação (eixo vertical à esquerda) e FP (eixo vertical à direita) são relativos a um mesmo conjunto de testes e aos classificadores treinados em 5 rodadas independentes sendo que os resultados mostrados equivalem à média dos resultados de todas as repetições. Apesar da presença de oscilação, uma clara tendência de alta na taxa de Recuperação é observada à medida que o conjunto de amostras positivas é ampliado artificialmente. O desvio padrão para a Recuperação ficou em torno de 3% para todos os valores de N exceto para N = 285 que apresentou maior variação entre rodadas distintas de treinamento.

Uma vez que cada conjunto de amostras positivas foi gerado de maneira aleatória e independente, não há garantias de que os conjuntos menores de amostras positivas estão contidos nos conjutos maiores e parte da oscilação nos resultados deve estar relacionada a esse fato. Esse experimento reforça a hipótese de que um conjunto de amostras positivas de treinamento relativamente pequeno pode ser usado para alcançar uma boa taxa de detecção no caso específico de FIPs.



Figura 5.6: Efeito do aumento de amostras positivas de treinamento, média de 5 rodadas.

# Capítulo 6

# **Resultados Experimentais**

Este capítulo apresenta os estudos realizados para a escolha de parâmetros de detecção nos dois estágios da abordagem proposta, uma análise dos impactos de suas variações e o desempenho final alcançado em termos de recuperação, precisão e tempo de processamento. Uma análise qualitativa dos problemas mais comuns nos resultados é apresentada destacando os desafios encontrados.

## 6.1 Conjuntos de Dados

A determinação dos parâmetros de treinamento do detector de FIPs descrita no capítulo 5 foi realizada com um conjunto de 380 FIPs disponível inicialmente.

Para a realização dos experimentos finais novas amostras foram coletadas em complemento a esse conjunto inicial e o conjunto total de amostras disponíveis ao final foi dividido em subconjuntos conforme descrição a seguir:

**Conjunto A:** Imagens de boa qualidade contendo apenas um código QR cada, totalizando 1005 FIPs recortados em 335 imagens.

**Conjunto B:** Imagens de qualidade ruim contendo apenas um código QR cada, totalizando 150 FIPs recortados em 50 imagens.

**Conjunto C:** Imagens de qualidade variada contendo mais de um código QR cada, totalizando 207 FIPs recortados em 25 imagens.

**Conjunto T:** Todas as imagens contendo códigos QR cujos FIPs foram anotados, ou seja,  $T = A \cup B \cup C$  totalizando 1362 FIPs recortados de 410 imagens.

**Conjunto S:** Imagens que não foram recortadas/anotadas para marcar a posição de FIPs, totalizando 100 imagens com 502 FIPs.

**Conjunto U:** Todas as imagens contendo códigos QR, ou seja,  $U = T \cup S$  totalizando 510 imagens com 1864 FIPs.

A determinação dos parâmetros de detecção de FIPs foi realizada a partir de dois subconjuntos de A, disjuntos entre si e denominados  $A_1$  e  $A_2$  conforme descrição a seguir:

**Conjunto**  $A_1$ : Subconjunto de A usado para treinamento do detector de FIPs, totalmente disjunto de  $A_2$ , formado por 154 imagens contendo apenas um código QR cada uma e um total de 462 FIPs.

**Conjunto**  $A_2$ : Subconjunto de A usado para o teste do detector de FIPs, totalmente disjunto de  $A_1$ , formado por 74 imagens contendo apenas um código QR cada uma e um total de 222 FIPs.

Para a determinação dos parâmetros de detecção de códigos QR completos, foi usado um subconjunto de A, disjunto de  $A_1$  e  $A_2$ , denominado  $A_3$  conforme descrição a seguir:

**Conjunto**  $A_3$ : Subconjunto de A usado para ajuste dos parâmetros de detecção do segundo estágio da abordagem proposta, composto por 45 imagens contendo apenas um código QR cada e um total de 135 FIPs.

Todas as imagens que compõem os conjuntos descritos foram obtidas através do sítio www.flickr.com.

## 6.2 Detecção de FIPs

## 6.2.1 Parâmetros de detecção

A aplicação de um detector baseado na técnica proposta por Viola-Jones depende de alguns parâmetros que influenciam significativamente seu desempenho, chamados de parâmetros de detecção, representados por SF e ND e descritos na seção 5.1.4.

Um detector de FIPs foi treinado usando o conjunto  $A_1$  de 462 amostras positivas (FIPs recortados) e 820 imagens negativas contendo grande variedade de cenários mas nenhum código QR ou FIP. Os parâmetros de treinamento usados estiveram de acordo com os determinados no capítulo 5.

O detector de FIP treinado foi avaliado através do conjunto de teste  $A_2$  consistindo de 74 imagens contendo 222 amostras positivas (exatamente um código QR completo em cada imagem), totalmente distinto do conjunto de treinamento. Durante essa avaliação, a influência dos parâmetros "SF" e "ND" (apresentados na seção 5.1.4) foi analisada e os resultados são apresentados a seguir.

**Fator de Escalonamento (SF):** A Figura 6.1 mostra que o fator de escalonamento (SF) impacta diretamente a qualidade dos resultados produzidos pelo detector. Os resultados

variam em uma larga faixa que vai desde SF=1.05 (recuperação: 0.98 / FP: 628) a SF=1.30 (recuperação: 0.86 / FP: 93). O aumento do fator de escalonamento acarreta na diminuição da recuperação e do número de FP. Considerando a existência do segundo estágio que deve ser capaz de lidar com FP mas não com FN, a taxa de recuperação foi priorizada e o fator de escalonamento escolhido foi de SF=1.10, de acordo com o destaque no gráfico (recuperação: 0.99 / FP: 290) da figura 6.1.



**Figura 6.1:** Recuperação e FP apresentados por um mesmo detector de FIP com variação do parâmetro de detecção SF (fator de escalonamento).

Número mínimo de detecções sobrepostas (ND): Conforme denifido na seção 5.1.4, esse parâmetro indica a quantidade mínima de detecções sobrepostas que devem existir em uma dada posição para que as mesmas, em conjunto, sejam consideradas uma única detecção. Na implementação usada, duas detecções são consideradas sobrepostas se sua posição em x e em y difere de menos de 20% da largura da menor das duas detecções e, entre si, a largura das duas detecções não difere em mais de 20%. A Figura 6.2 mostra uma curva ROC comparando a taxa de recuperação e o número de FP enquanto ND varia de 1 a 200. À primeira vista uma boa escolha de ND seria 53 conforme destacado no ponto "A" da Figura 6.2. Contudo, ao considerar que esses resultados correspondem ao primeiro estágio da abordagem proposta e dizem respeito apenas à detecção de FIPs, a escolha de ND deve priorizar nesse estágio a recuperação com grande ênfase. Isso se deve ao fato de que o segundo estágio é capaz de eliminar a maioria dos FP sem, por outro lado, ter a possibilidade de detectar um código QR onde um de seus FIPs tenha deixado de ser encontrado no primeiro estágio, ou seja, um

FN do primeiro estágio. Dessa forma o valor escolhido foi de ND=30 conforme o ponto "B" destacado na Figura 6.2.



**Figura 6.2:** Recuperação  $\times$  FP para o detector de FIPs treinado com variação do parâmetro de detecção ND (número mínimo de detecções sobrepostas) de 1 a 200.

Com os parâmetros determinados (SF=1.10 e ND=30) o detector de FIP alcançou uma taxa de recuperação de 0.986 (detectando 219 dos 222 FIPs) e FP de 284 (em 74 imagens de teste) no conjunto de teste. Essa taxa de recuperação bastante alta deve-se em parte ao fato de que os parâmetros SF e ND foram ajustados priorizando-os nesse mesmo conjunto de teste.

## 6.3 Detecção de Códigos QR

## 6.3.1 Parâmetros de detecção

O algoritmo 1 que representa o segundo estágio da abordagem proposta foi desenvolvido de maneira a permitir um ajuste fino das restrições geométricas por ele impostas. Os parâmetros usados para tal ajuste,  $d \in \varepsilon$  (descritos na seção 4.3.1), alteram o rigor dos critérios de geometria e tamanho respectivamente. Dessa forma é possível ajustar independentemente o rigor das restrições nesses dois aspectos da geometria do arranjo de candidatos a FIP. Variações nesses parâmetros foram avaliadas de forma a determinar a configuração de melhor desempenho nesse conjunto de validação.

Um conjunto de validação, chamado de conjunto  $A_3$ , que consiste de 45 imagens contendo exatamente um código QR completo foi usado exclusivamente para o ajuste de parâmetros de detecção para o estágio de agregação de candidatos a FIP. O detector de FIP descrito na seção 6.2.1 foi aplicado juntamente com os parâmetros SF = 1.10 e ND = 30.

Por fim, uma análise do número mínimo de detecções sobrepostas (ND) foi feita sob a ótica do reconhecimento de códigos QR completos e não apenas FIPs de forma a consolidar o conjunto de parâmetros completo a ser utilizado pela abordagem proposta.

È importante salientar que nesta seção os valores de recuperação e FP dizem respeito a códigos QR completos e não a FIPs isolados.

### 6.3.2 Tolerância dos critérios de restrições geométricas

Os parâmetros do segundo estágio (agregação de candidatos a FIP) consistem nas tolerâncias  $\varepsilon$  e *d* definidas nos critérios de tamanho e geometria, respectivamente, e tiveram seus valores variados durante a aplicação de um mesmo detector de FIP para determinar a melhor configuração no conjunto de validação.

A Figura 6.3 mostra as taxas de recuperação e o número de FP que foram alcançados para diferentes valores do parâmetro de tolerância do critério de geometria (d) no conjunto de validação  $A_3$  relativos à detecção de códigos QR completos. Uma tolerância demasiadamente pequena (d = 0.05) resulta em uma baixa taxa de recuperação. Com o aumento da tolerância a recuperação aumenta naturalmente, mas é acompanhada pelo aumento de FP. O valor destacado (d = 0.25) corresponde à escolha feita para o parâmetro de tolerância do critério de geometria que, em graus, equivale a uma tolerância de aproximadamente 14.5°. Detalhes sobre esse parâmetro encontram-se na seção 4.3.1.

A Figura 6.4 mostra a recuperação e FP para códigos QR completos usando o conjunto de validação  $A_3$  e fixando a tolerância do critério de geometria em d = 0.25. O valor destacado representa a escolha feita para o parâmetro de tolerância do critério de tamanho  $\varepsilon = 0.40$ , detalhado na seção 4.3.1. Do mesmo modo que ocorre para a tolerância do critério de geometria, valores maiores para esse parâmetro aumentam a taxa de recuperação mas consequentemente aumentam o FP.



**Figura 6.3:** Recuperação e FP do detector de códigos QR completos (estágios 1 e 2) sujeitos à variação do parâmetro de tolerância do critério de geometria d com valor fixo para o parâmetro de tolerância do critério de tamanho  $\varepsilon = 0.30$ .

# 6.3.3 Número mínimo de detecções sobrepostas em códigos QR completos

A Figura 6.5 mostra a recuperação e FP do detector de FIP aplicado ao conjunto de validação  $A_3$  ao variar apenas o parâmetro de detecção ND (número mínimo de detecções sobrepostas) e manter fixos  $\varepsilon = 0.4$  e d = 0.25 conforme determinados na seção 6.3.2. O valor destacado ND = 30 foi o escolhido como ajuste final dos parâmetros de detecção para códigos QR completos.



**Figura 6.4:** Recuperação e FP apresentados pelo detector de códigos QR completos (estágios 1 e 2) enquanto é sujeito à variação do parâmetro de tolerância do critério de tamanho  $\varepsilon$  com valor fixo para o parâmetro de tolerância do critério de geometria d = 0.25.



**Figura 6.5:** Recuperação e FP do detector de códigos QR completos (estágios 1 e 2) com valores fixos dos parâmetros  $\varepsilon = 0.4$  e d = 0.25 sujeitos à variação do número mínimo de detecções sobrepostas (ND).

## 6.4 Avaliação do Detector de Códigos QR

A avaliação quantitativa do desempenho do detector de códigos QR foi realizada separadamente para os estágios 1 e 2 da abordagem proposta. Foram realizadas duas rodadas de validação cruzada além de testes sobre conjuntos particulares de amostras que exibiam, por exemplo, múltiplos códigos QR em cada imagem ou baixa qualidade. Qualitativamente foram analisados os problemas mais comuns e interessantes durante a detecção de códigos QR onde são evidenciadas qualidades e limitações da abordagem proposta.

## 6.4.1 Avaliação Quantitativa

#### Validação Cruzada

O desempenho de duas cascatas de classificadores treinadas independentemente através do uso dos mesmos parâmetros pela ferramenta opencv-haartraining pode variar substancialmente em razão da influência das amostras positivas e negativas utilizadas. Tal fato ocorre em função da escolha aleatória de amostras negativas que é feita antes do treinamento de cada estágio da cascata.

Para garantir a consistência dos resultados apresentados, foram realizadas duas rodadas de validação cruzada nas quais o conjunto de imagens utilizado foi dividido em 5 e foram realizadas 5 etapas durante as quais 4/5 das imagens foram utilizados para treinamento do detector de FIPs e 1/5 para teste. Em resumo, as rodadas podem ser descritas como se segue:

Rodada 1 O conjunto "A" (amostras de boa qualidade) foi dividido em 5 subconjuntos onde 4 foram usados para treinamento e 1 para testes em cada etapa.

**Rodada 2** O conjunto formado por amostras de " $A \cup B$ " (amostras de qualidade variada) foi dividido em 5 subconjuntos onde 4 foram usados para treinamento e 1 para testes em cada etapa.

A seguir são apresentados os resultados das duas rodadas de validação cruzada em termos de desempenho da detecção de FIPs e de códigos QR completos, respectivamente.

**Estágio 1 - Detecção de FIPs** Os parâmetros de treinamento dos classificadores foram determinados com base na análise em 5.2.2. Os valores abaixo permaneceram idênticos aos determinados em 5.2.2:

- Conjunto de Características (MODE): Básico;
- Simetria (SYM): Simétrico;
- Topologia do Classificador (MTS): Cascata;
- Número de divisões do classificador fraco (NS): 1;

- Máxima taxa de alarme falso (FA): 0.50;
- Tamanho das amostras de treinamento:  $16 \times 16$ .

Além disso, foram usados os parâmetros abaixo:

- Mínima taxa de acerto (HR): 0.98;
- Número de estágios da cascata (ST): 19;
- Número de amostras de treinamento na rodada 1 (SAMPLES): 8040;
- Número de amostras de treinamento na rodada 2 (SAMPLES): 9240;

Os parâmetros HR e ST não foram avaliados durante a determinação de parâmetros pois devem ser ajustados em função do número de estágios esperados na cascata resultante e na taxa mínima esperada de recuperação das amostras analisadas. HR é um limitante inferior para a recuperação de amostras positivas em cada estágio da cascata de classificadores durante o treinamento sendo que, dessa forma, o parâmetro HR=0.98 limita o descarte de amostras positivas (como falsos negativos) a 2% para o conjunto de treinamento. O número de estágio da cascata (ST=19) combinado com a mínima recuperação (HR=0.98) e máxima taxa de alarmes falsos (FA=0.50) resulta em uma cascata de classificadores que apresenta, relativamente ao conjunto de treinamento, taxa de alarmes falsos menor que  $0.5^{19} = 1.9 \times 10^{-6}$  e taxa de recuperação maior ou igual a  $0.98^{19} = 0.68$ . O treinamento do classificador é interrompido quando a taxa de alarmes falsos durante o treinamento de algum estágio é inferior à taxa esperada ao fim do treinamento ou quando o número de estágios é atingido.

O número de amostras de treinamento (SAMPLES) foi aumentado de 4000 (valor utilizado em 5.2.2) para 8040 na rodada 1 e 9240 na rodada 2 como forma de manter aproximadamente a mesma razão entre amostras originais e as criadas artificialmente. Durante a determinação dos parâmetros de treinamento foram utilizadas 380 amostras de FIPs originais e esse conjunto foi expandido para 4000, enquanto nas rodadas de validação cruzada foram utilizadas 804 e 924 amostras originais de FIPs ampliando-as em um fator de 10 vezes para 8040 e 9240 nas rodadas 1 e 2, respectivamente.

Etapa	TP	FN	FP	Recuperação	Precisão
Etapa 1	196	5	104	0,9751244	0,6533333
Etapa $2$	192	9	163	0,9552239	0,5408451
Etapa 3	197	4	124	$0,\!9800995$	$0,\!6137072$
Etapa 4	192	9	107	$0,\!9552239$	$0,\!6421405$
Etapa 5	197	4	74	$0,\!9800995$	0,7269373
Média	$194,\!8$	$^{6,2}$	$114,\!4$	0,9691542	$0,\!6353927$

Tabela 6.1: Resultados para a detecção de FIPs - Rodada 1.

Etapa	TP	FN	FP	Recuperação	Precisão
Etapa 1	221	10	110	0,95671	0,6676737
Etapa 2	223	8	143	0,965368	0,6092896
Etapa 3	224	7	96	0,969697	0,7
Etapa 4	226	5	125	$0,\!978355$	0,6438746
Etapa 5	229	2	217	0,991342	0,5134529
Média	$224,\!6$	$^{6,4}$	138,2	0,9722944	0,6268582

 Tabela 6.2: Resultados para a detecção de FIPs - Rodada 2.

A recuperação e precisão médias alcançadas na detecção de candidatos a FIP na rodada 1 de validação cruzada foram, respectivamente, 0,97 e 0,64 conforme a Tabela 6.1.

A recuperação e precisão médias alcançadas na detecção de candidatos a FIP na rodada 2 de validação cruzada foram, respectivamente, 0, 97 e 0, 63 conforme a Tabela 6.2.

Nota-se que a diferença de desempenho entre as rodadas de validação cruzada foi muito pequena, sugerindo que a adição de 15% de amostras de qualidade ruim (conjunto "B") ao conjunto "A" não teve impactos negativos. Ao considerar que em média os testes realizados na rodada 2 são mais difíceis por conterem as imagens de baixa qualidade provenientes do conjunto "B" os resultados indicam que o uso de amostras de qualidade variada é benéfico.

Estágios 1 e 2 - Detecção de códigos QR completos Os resultados de detecção de códigos QR completos foram produzidos em validação cruzada conforme descrição feita acima. Cada detector de FIPs (estágio 1) treinado com 4/5 das imagens disponíveis foi aplicado às imagens restantes (1/5) e seus resultados foram processados pelo algoritmo de agregação de candidatos a FIP (estágio 2) utilizando os parâmetros de detecção determinados em 6.2.1 e 6.3.1. Eventualmente um mesmo FIP pode ser detectado como dois candidatos a FIPs distintos cujos centros estão posicionados próximos à localização exata do centro do FIP na imagem. Isso pode fazer com que o estágio 2 aponte a detecção de dois códigos QR que, apesar de serem distintos por serem compostos de trios distintos de candidatos a FIPs detectados no estágio 1, correspondem na realidade ao mesmo código QR na imagem. Esse tipo de situação configura as detecções repetidas apontadas na coluna "Repetidas" da Tabela 6.3 e da Tabela 6.4.

- Fator de Escalonamento (SF, Estágio 1) = 1.10;
- Número mínimo de detecções sobrepostas (ND, Estágio 1) = 30;
- Menor tamanho de detecção =  $16 \times 16$ ;
- Tolerância do critério de tamanho ( $\varepsilon$ , Estágio 2) = 0.40;
- Tolerância do critério de geometria (d, Estágio 2) = 0.25;

A recuperação e precisão médias alcançadas para a detecção de códigos QR durante a rodada 1 de validação cruzada foram, respectivamente, 0,94 e 0,76 conforme a Tabela 6.3.

Etapa	TP	FN	FP	Repetidas	Recuperação	Precisão
Etapa 1	64	3	61	0	0,9552239	0,512
Etapa 2	62	5	23	0	0,9253731	0,729411765
Etapa 3	64	3	17	1	0,9552239	0,790123457
Etapa 4	60	7	9	0	$0,\!8955224$	0,869565217
Etapa 5	65	2	7	2	$0,\!9701493$	0,902777778
Média	63	4	$23,\!4$	0,6	0,9402985	0,760775643

 Tabela 6.3: Resultados para a detecção de códigos QR completos - Rodada 1.

Cabe ressaltar que a etapa 1 apresentou um caso muito ruim representado por uma única imagem na qual houveram 52 detecções falsas de códigos QR, de um total de 61 FP dessa rodada. Algo similar ocorreu na etapa 3 onde uma única imagem foi responsável por 10 dos 17 alarmes falsos. Essas imagens ilustram um dos tipos de erro difíceis de combater através da abordagem proposta pois ocorrem devido a um número elevado de falsos positivos de tamanho e distância compatíveis com os critérios de restrições geométricas estabelecidos. A seção 6.4.2 descreve qualitativamente os principais tipos de erros encontrados durante os experimentos.

Etapa	TP	FN	FP	Repetidas	Recuperação	Precisão
Etapa 1	68	9	15	0	0,8831169	0,819277108
Etapa 2	69	8	17	0	$0,\!8961039$	0,802325581
Etapa 3	73	4	8	1	0,9480519	0,901234568
Etapa 4	71	6	21	0	0,9220779	0,77173913
Etapa 5	72	5	61	0	$0,\!9350649$	0,541353383
Média	$70,\!6$	6,4	24,4	0,2	0,9168831	0,767185954

Tabela 6.4: Resultados para a detecção de códigos QR completos - Rodada 2.

A recuperação e precisão médias alcançadas para a detecção de códigos QR durante a rodada 2 de validação cruzada foram, respectivamente, 0,92 e 0,77 conforme a Tabela 6.4. Nota-se uma queda pequena na taxa de recuperação da segunda rodada. Considerando que durante a rodada 1 de validação cruzada a taxa de recuperação de FIPs foi a mesma, pode-se considerar como um dos fatores que influenciaram a queda da taxa de recuperação na rodada 2 a existência de um número maior de imagens com distorções perspectivas ou de superfície. Esse tipo de imagem, considerada em detalhes na seção 6.4.2, tende a provocar falhas na detecção apenas do segundo estágio da abordagem proposta, ou seja, os três FIPs de um mesmo código QR são detectados corretamente mas por terem tamanhos relativamente diferentes e terem suas distâncias e ângulos distorcidos, não são considerados um código QR.

#### Desempenho Final do Detector de Códigos QR

Um classificador denominado de "final" foi treinado com todas as imagens do conjunto A e utilizado para avaliar o desempenho da abordagem proposta em situações de baixa qualidade, múltiplos códigos QR em uma mesma cena e situações variadas. Para o treinamento, as 1005 amostras do conjunto A foram submetidas a variações aleatórias, conforme procedimento descrito em 5.2.3, aumentado a quantidade de amostras disponíveis para treinamento em 10 vezes.

Os parâmetros usados no treinamento foram os seguintes:

- Conjunto de Características (MODE): Básico;
- Simetria (SYM): Simétrico;
- Topologia do Classificador (MTS): Cascata;
- Número de divisões do classificador fraco (NS): 1;
- Máxima taxa de alarme falso (FA): 0.50;
- Tamanho das amostras de treinamento:  $16 \times 16$ .
- Mínima taxa de acerto (HR) : 0.98;
- Número de estágios da cascata (ST): 19;
- Número de amostras de treinamento (SAMPLES): 10050;

Para detecção, os parâmetros utilizados foram:

- Fator de Escalonamento (SF, Estágio 1) = 1.10;
- Número mínimo de detecções sobrepostas (ND, Estágio 1) = 30;
- Menor tamanho de detecção =  $16 \times 16$ ;
- Tolerância do critério de tamanho ( $\varepsilon$ , Estágio 2) = 0.40;
- Tolerância do critério de geometria (d, Estágio 2) = 0.25;

A tabela 6.5 apresenta o desempenho para detecção de FIPs do classificador "final".

Tabela 6.5: Resultados para a detecção de FIPs

Conjunto	TP	FN	$\mathbf{FP}$	Recuperação	Precisão
В	127	26	42	$0,\!830065359$	0,75147929
C	200	10	90	0,952380952	0,689655172
S	345	157	244	$0,\!687250996$	0,58573854

A tabela 6.6 apresenta o desempenho para detecção de códigos QR completos do classificador "final".

Nota-se que o desempenho nos conjuntos "B", "C" e "S" foi consideravelmente inferior ao alcançado durante as rodadas de validação cruzada em termos de recuperação. Em regra,

Conjunto	TP	FN	FP	Recuperação	Precisão
В	33	18	5	$0,\!647058824$	0,868421053
$\mathbf{C}$	65	6	170	0,915492958	$0,\!276595745$
$\mathbf{S}$	78	77	62	0,503225806	0,557142857

 Tabela 6.6:
 Resultados para a detecção de códigos QR completos

todas as imagens marcadas, contidas nos conjuntos "A", "B" e "C", foram escolhidas para marcação por apresentarem códigos com relativa clareza enquanto as imagens de "S" em sua maioria não foram marcadas justamente por apresentarem qualidade baixa. Isso explica em parte a grande diferença de desempenho entre os resultados em "S" e os resultados sobre os outros conjuntos. Outra observação importante é que a precisão para detecção de códigos QR completos no conjunto D é baixa, como devia-se esperar por serem imagens onde existe mais de um código QR e, portanto, mais sujeitas à formação de falsos positivos por interferência entre os símbolos.

## 6.4.2 Avaliação Qualitativa

Após a determinação dos parâmetros de treinamento e de detecção para os dois estágios da abordagem proposta e das medições de desempenho dos dois estágios através de validação cruzada, um detector final de FIPs foi produzido. Esse detector foi treinado com os parâmetros de treinamento determinados experimentalmente e com o conjunto total de amostras de FIPs disponíveis. Dessa forma foi produzida a implementação final da abordagem proposta que pode ser usada em aplicações futuras.

Um conjunto de imagens, totalmente disjunto dos conjuntos usados até então para treinamento do detector de FIPs e ajuste dos parâmetros, foi processado e os resultados ilustram por um lado vários casos onde a abordagem proposta é eficaz e por outro casos que exaltam suas limitações. Seguem exemplos de imagens processadas por essa implementação final da abordagem proposta.

Exemplos adicionais de imagens processadas podem ser encontradas a partir do endereço http://www.vision.ime.usp.br/demos.

#### Tipos de falhas na detecção de códigos QR completos

A grande maioria dos casos em que a abordagem proposta falha na detecção de códigos QR, seja por apontar falsos positivos ou falsos negativos, pode ser classificada em alguma das categorias abaixo relacionadas. De maneiras diferentes, cada uma das categorias de erros apresentadas revela limitações do primeiro ou segundo estágio da abordagem proposta e devem sempre ser analisados levando em consideração a probabilidade de sua ocorrência em casos gerais e comuns de uso. Presença de múltiplos códigos QR próximos na imagem: A presença de múltiplos códigos QR dispostos em uma mesma região da imagem é difícil de tratar através da abordagem proposta pois nessa situação três FIPs que não sejam parte do mesmo símbolo facilmente podem acabar por representar um arranjo que satisfaça as restrições geométricas impostas, como pode ser observado na figura 6.6. Esse tipo de imagem ocorre apenas em situações específicas e que não são muito frequentes nos casos gerais. Esse tipo de erro pode ser bastante diminuído caso se opte por restringir a versão máxima de código QR a ser detectado (nesse trabalho a versão não foi limitada, portanto códigos de versões 1 a 40 podem ser detectados).



Figura 6.6: Múltiplos códigos QR próximos na imagem (www.flickr.com, blech).

**Presença de um falso positivo para FIP próximo a um código QR:** A presença de um falso positivo para FIP próximo a um código QR tem grande probabilidade de ocasionar um falso positivo para a detecção de código QR como exemplificado na figura 6.7. Dos erros comuns na detecção de códigos QR esse é o que ocorre com mais frequência mas que tem o menor impacto negativo em termos de localização do código pois o falso positivo gerado contempla parcialmente um código QR verdadeiro e, portanto, encontra-se em uma região próxima ao símbolo que de fato está presente na cena.

Falsos negativos na detecção de FIPs: Esse pode ser considerado o erro de maior impacto nos resultados em geral pois, apesar de ocorrer moderadamente, a abordagem proposta



Figura 6.7: Duas detecções corretas e um falso positivo. (www.flickr.com scott\_blake)

em sua formulação atual não é capaz de superar um falso negativo para FIP, no primeiro estágio, conforme exemplificado na figura 6.8. Muitos são os fatores que podem induzir a um falso negativo no primeiro estágio, como variações extremas de iluminação, ruído, oclusão parcial, escala reduzida ou inclinação. A figura 6.9 mostra um caso peculiar onde tem-se uma detecção bem sucedida apesar da oclusão parcial de um FIP e um falso negativo para a detecção do segundo código QR em virtude de um FIP não detectado. Destaque-se que a inclinação do código não é fator determinante para o insucesso na detecção de FIPs uma

vez que dos três FIPs igualmente inclinados apenas um não foi detectado. Observa-se que a inclinação, bem como os outros fatores citados, apenas diminuem de maneira estatística a chance de detecção correta de FIPs.



**Figura 6.8:** Uma detecção correta e um falso negativo devido a um FIP não detectado. (www.flickr.com davidking)

**Distorção perspectiva acentuada em símbolos pequenos:** A aproximação das distâncias e dimensões de FIPs feitas com base em pixels para cálculo das restrições geométricas em símbolos de pequenas dimensões tende a introduzir erros de arredondamento nos cálculos que podem alterar, de modo representativo, o significado geométrico das restrições impostas no segundo estágio. Sendo assim, símbolos pequenos podem vir a ser julgados incompatíveis com as restrições do segundo estágio mesmo não o sendo, como é o caso dos dois códigos QR à esquerda na Figura 6.10 que contém, ao todo, 3 falsos negativos.

**Presença de falsos positivos para FIP arranjados regularmente:** Regiões de texto ou que exibem algum padrão de espaçamento regular constituem um caso de particular sensibilidade para a abordagem proposta, como o mostrado na Figura 6.11. Uma vez que alguns desses padrões repetitivos sejam identificados incorretamente pelo primeiro estágio como FIPs existe uma boa chance de que eles estejam também alinhados de forma regular o que aumenta a probabilidade de erro. Casos como esses são raros mas, quando ocorrem, podem resultar em um grande número de alarmes falsos em um mesmo quadro.



**Figura 6.9:** Uma detecção com oclusão parcial de um FIP e um falso negativo, possivelmente devido à inclinação. (www.flickr.com cjsveningsson)



Figura 6.10: Três falsos negativos. (www.flickr.com davidking)



**Figura 6.11:** Uma detecção correta e outra formada por três falsos positivos para FIP. (www.flickr.com, qriouscode)

**Casos de sucesso:** Em geral a abordagem proposta foi capaz de obter sucesso em uma grande variedade de condições conforme ilustrado pelos exemplos nas Figuras 6.12 a 6.16.



Figura 6.12: Exemplo de detecção correta de um código QR (www.flickr.com gaku).



**Figura 6.13:** Exemplo de detecção correta de um código QR com vários falsos positivos para FIPs (www.flickr.com whoisstan).


**Figura 6.14:** Exemplo de detecção correta de um código QR com vários falsos positivos para FIPs (www.flickr.com whoisstan).



Figura 6.15: Exemplo de detecção correta de um código QR (www.flickr.com whoisstan).



**Figura 6.16:** Exemplo de detecção correta de um código QR com dois falsos positivos para FIPs (www.flickr.com whoisstan).

#### 6.5 Processamento de Vídeo

As possibilidates mais interessantes de aplicação da abordagem proposta estão relacionadas ao processamento de vídeo e, portanto, é crucial que se consiga processar quadros de um fluxo de vídeo em tempo razoavelmente pequeno. Observa-se que o tempo despendido no primeiro estágio (detecção de FIPs) é muito superior ao despendido no segundo estágio (agregação de candidatos a FIP) e torna-se portanto determinante do tempo total de procesamento. Além disso, fica evidente que o tempo necessário para processamento do primeiro estágio é grandemente influenciado pelo fator de escalonamento (SF) e o menor tamanho de detecção. Os resultados nesta seção revelam a influência desses dois parâmetros no desempenho computacional da abordagem proposta e ressaltam a flexibilidade proporcionada por eles, uma vez que o compromisso entre qualidade dos resultados e a velocidade de processamento dos quadros pode ser facilmente gerenciado através de seu ajuste. Tal flexibilidade é valiosa quando se considera que possivelmente as implementações da abordagem proposta serão executadas em uma grande variedade de dispositivos.

Os experimentos a seguir foram executados em uma máquina virtual sobre um computador com um processador de dois núcleos e os resultados apresentados correspondem a valores médios obtidos em 1000 quadros. O vídeo é um fluxo de 640x480 pixels adquirido através de uma câmera simples onde dois códigos QR são movimentados pelo laboratório variando distância e ângulo de inclinação de acordo com um roteiro fixo e pré-definido. A Figura 6.17 mostra a janela de execução durante o processamento de um fluxo de vídeo.

### 6.5.1 Tamanho Mínimo de Detecção de FIPs no Processamento de Vídeo

O tamanho do menor FIP que pode ser encontrado no primeiro estágio está limitado inferiormente pelo tamanho das amostras utilizadas durante o treinamento do detector. Por construção, o detector trabalha com imagens de tamanho fixo determinado no momento de seu treinamento sendo que amostras maiores são reduzidas a esse tamanho antes de serem submetidas à cascada de calssificadores.

Se existe o conhecimento prévio de que todos os FIPs que se pretende detectar são maiores que um dado tamanho mínimo (maior que o tamanho das amostras usadas para treinar o detector) não é necessário considerar todas as amostras possíveis mas sim apenas as de tamanho maior ou igual a tal tamanho mínimo. Essa medida pode reduzir consideravelmente o tempo de processamento de uma imagem completa pois pode reduzir em muito o espaço de busca a ser explorado. A figura 6.18 mostra a variação no tempo de processamento quando esse tamanho mínimo de detecção é variado para um fator de escalonamento fixo em SF=1.20. O tempo de processamento dos quadros varia de 22 ms a 103 ms. Note que o tempo de execução do segundo estágio (agregação de candidatos a FIP), mostrado no eixo direito do gráfico, tem pouca contribuição no tempo total de processamento do quadro



Figura 6.17: Quadro retirado de um dos fluxos de vídeo experimentais.

correspondendo aproximadamente a apenas 0.8% do tempo de processamento do primeiro estágio (detecção de candidados a FIP).

#### 6.5.2 Fator de Escalonamento no Processamento de Vídeo

Outro parâmetro crítico no desempenho computacional da abordagem proposta é o fator de escalonamento (SF) a ser usado durante o primeiro estágio (detecção de candidatos a FIP). Quanto menor o fator de escalonamento, mais lentamente a janela deslizante de aquisição de amostras é aumentada e maior é o número total de amostras a serem processadas. Para avaliar os efeitos do fator de escalonamento no tempo de processamento, o menor tamanho de detecção fixado nos testes foi o menor possível, ou seja, o tamanho usado para treinar o detector de FIPs que corresponde a  $16 \times 16$  pixels. O fator de escalonamento (SF) foi variado de 1.05 a 1.30 e a figura 6.19 mostra que os tempos de processamento foram de 45 ms a 255 ms por quadro equivalendo a uma taxa de 22.0 e 3.9 quadros por segundo respectivamente.

Exemplo de vídeo processado pode ser encontrado em http://www.vision.ime.usp.br/ demos.



**Figura 6.18:** Tempo de processamento dos estágios 1 e 2 (resultados em ms, fluxo de vídeo de  $640 \times 480$  pixels, média de 1000 quadros).

### 6.6 Discussão

Os resultados apresentados indicam que a técnica de detecção rápida de objetos proposta por Viola-Jones pode ser empregada com sucesso à detecção de códigos QR. A ausência de estrutura rígida desses códigos de barras pode ser compensada por uma abordagem de detecção por componentes (FIPs) seguida de um pós-processamento que agrega os resultados.

A escolha dos melhores parâmetros foi feita tomando como base não só recuperação e FP (falsos positivos) pois grande importância foi dada também à simplicidade do detector resultante. Além disso, especial atenção foi dada à recuperação pois o algoritmo de pósprocessamento não é capaz de emitir uma detecção válida caso qualquer um dos três FIPs não tenha sido detectado. Dessa forma é preferível que a etapa de detecção retorne um número elevado de falsos positivos a deixar de detectar algum FIP.

O balanceamento da cascata gerada (uniformidade na quantidade de características avaliadas entre estágios distintos) também foi considerado dado que essa característica desejável ressalta a capacidade de foco seletivo do processamento que a própria abordagem em cascata oferece.

O aumento artifical das amostras de treinamento mostrou-se eficaz e foi utilizado para a produção do detector final.

Apesar de Lienhart *et al.* (2002) destacar a utilidade do conjunto extendido de *haar-like features* e do estabelecimento de classificadores fracos como pequenas árvores no domínio de detecção de faces, essas observações não se repetiram nos experimentos desse trabalho para



**Figura 6.19:** Tempo de processamento para os estágios 1 e 2 (resultados em ms, fluxo de vídeo de 640 × 480 pixels, valores médios para 1000 quadros).

determinação dos parâmetros de treinamento.

Em relação a desempenho, constatou-se que o tempo de pós-processamento é quase negligível em comparação ao tempo de detecção e que ambos podem ser executados em tempo real, processando entre 3.9 e 22 quadros por segundo em laboratório. Faz-se a ressalva de que essa relação entre tempo de detecção e de pós-processamento está condicionada à existência de um número baixo de alarmes falsos (nos testes  $FA \approx 3 \times Hits$ ).

# Capítulo 7

### Conclusão e Trabalho Futuro

Esse trabalho abordou o problema de detecção de códigos QR em imagens adquiridas ao acaso, um passo necessário para viabilizar a aplicação dessa tecnologia de código de barras em situações onde não se pode presumir a existência de um operador de câmera capaz de enquadrar corretamente o símbolo de interesse.

Para resolver esse problema desafiador que requer simultaneamente processamento em tempo real e robustez em relação a escala, iluminação, rotação e distorções perspectivas, foi proposta uma abordagem de detecção composta por dois estágios e baseada em componentes. Para o primeiro estágio foi proposto a detecção de partes visualmente destacadas dos códigos QR, conhecidas como padrões de posição ou FIPs, por meio do arcabouço para detecção rápida de objetos de Viola-Jones. A adaptação dos parâmetros de treinamento e detecção desse arcabouço para a detecção de FIPs foi feita com base em experimentos que revelaram possibilidades de ajustes finos de desempenho desse primeiro estágio. Para o segundo estágio foi proposto um algoritmo simples de agregação de candidatos a FIP que examina conjuntos de três deles para decidir, a partir da aplicação de restrições geométricas ao arranjo formado por esse conjunto, se eles indicam ou não a presença de um código QR naquela posição.

Resultados preliminares foram apresentados no artigo intitulado "Fast QR Code Detection in Arbitrary Acquired Images" publicado no XXIV SIBGRAPI (Belussi e Hirata 2011). Esse trabalho foi escolhido para uma edição especial do Journal of Mathematical Imaging and Vision e uma versão estendida do mesmo foi aceita para publicação sob o título de "Fast component-based QR code detection in arbitrarily acquired images".

A abordagem proposta foi validada e testada em um conjunto de imagens estáticas e mostrou desempenho satisfatório na detecção de códigos QR nelas presentes. Parte dos casos em que os códigos QR não foram detectados deve-se à não detecção de um ou mais FIPs de forma que, por construção do algoritmo de agregação de candidatos a FIP, um código QR não é detectado se algum de seus FIPs não tiver sido encontrado pelo primeiro estágio. Outra parte dos falsos negativos ocorre em virtude de as distorções geométricas presentes nos códigos QR serem maiores que as tolerâncias permitidas pelos parâmetros do segundo estágio. Quanto ao desempenho computacional, os testes em processamento de vídeo mostraram que a abordagem proposta foi capaz de fornecer respostas em tempo razoável e assim atingir uma taxa aceitável de quadros processados por segundo, compatível com as aplicações consideradas de tempo real.

O treinamento do detector de FIPs foi realizado ao final desse trabalho com o uso de 1005 amostras originais (ampliadas para 10050 através de variações aleatórias de iluminação e rotação em  $x, y \in z$ ). Apesar de ser um número expressivo de amostras, é relativamente baixo se comparado a tamanhos consagrados de conjuntos de treinamento para outros objetos mais complexos como faces onde se utiliza de 4000 a 10000 imagens originais. De qualquer forma, as taxas de recuperação de FIPs foram boas o suficiente para alcançar a qualidade relatada na detecção final. Conforme definido na seção 2.4, o presente trabalho restringe os efeitos de rotação e distorção perspectiva a níveis reduzidos. Essa decisão contribui para os bons resultados da detecção de FIPs mas certamente há espaço para melhorar o primeiro estágio da abordagem proposta ao buscar uma maneira de produzir um detector capaz de apresentar resultados comparáveis aos que foram obtidos sem excluir da análise códigos QR fortemente rotacionados ou distorcidos.

Como trabalhos futuros, a melhoria citada visando a robustez à rotação na detecção do primeiro estágio merece destaque. Quanto ao segundo estágio, o algoritmo de agregação de candidatos a FIP pode ser refinado de maneira a considerar em seu modelo a distorção perspectiva em função da diferença de tamanho entre os candidatos a FIP e então promover as devidas compensações. Outra linha de investigação para refinamento do segundo estágio seria estudar os histogramas ou mesmo os padrões na área interna do triângulo formado por um grupo de 3 candidatos a FIP que respeitem os critérios geométricos ou ao longo das linhas que unem os centros de dois candidatos a FIP. Essa segunda possibilidade teria ainda a vantagem de abrir caminho para a detecção de um código QR mesmo quando há um falso negativo de FIP no primeiro estágio. Por fim, pode ainda ser considerada a aplicação do detector de FIPs com parâmetros mais permissivos (por exemplo, diminuindo ND ou SF 5.1.4) nas seis regiões em que há maior probabilidade condicional de ocorrência de um FIP dado que dois candidatos a FIP já foram encontrados, como ilustrado na Figura 4.2.

O processamento de vídeo foi investigado apenas do ponto de vista de tempo de processamento dos quadros sendo que cada quadro foi tratado de maneira independente, como uma imagem totalmente distinta. Uma forma de reforçar os resultados, como trabalho futuro, é explorar a coerência temporal entre quadros sucessivos.

Os resultados apresentados em conjunto com as possibilidades de aperfeiçoamento enumeradas levam à conclusão de que a abordagem proposta tem grande potencial para viabilizar uma gama de aplicações inovadoras e importantes que dependem da detecção de códigos QR em imagens com enquadramento arbitrário.

# **Referências Bibliográficas**

- Agarwal et al.(2004) Shivani Agarwal, Aatif Awan e Dan Roth. Learning to detect objects in images via a sparse, part-based representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26:1475–1490. Citado na pág. 20
- Al-Khalifa(2008) Hend S. Al-Khalifa. Utilizing QR Code and Mobile Phones for Blinds and Visually Impaired People. Em Proceedings of the 11th International Conference on Computers Helping People with Special Needs, páginas 1065–1069. Springer-Verlag. Citado na pág. 2, 8, 9
- Belussi e Hirata(2011) L.F.F. Belussi e N.S.T. Hirata. Fast qr code detection in arbitrarily acquired images. Em *Graphics, Patterns and Images (Sibgrapi), 2011 24th SIBGRAPI Conference on*, páginas 281–288. doi: 10.1109/SIBGRAPI.2011.16. Citado na pág. 4, 30, 67
- Belussi e Hirata(2012) LuizF.F. Belussi e NinaS.T. Hirata. Fast component-based qr code detection in arbitrarily acquired images. Journal of Mathematical Imaging and Vision, páginas 1–16. ISSN 0924-9907. doi: 10.1007/s10851-012-0355-x. URL http://dx.doi.org/ 10.1007/s10851-012-0355-x. Citado na pág. 4
- Bradski(2000) G. Bradski. The OpenCV Library. Dr. Dobb's Journal of Software Tools. Citado na pág. 27
- Breiman et al. (1984) L. Breiman, J. H. Friedman, R. A. Olshen e C. J. Stone. Classification and Regression Trees. Wadsworth International Group. Citado na pág. 11, 14
- Canadi et al.(2010) Michael Canadi, Wolfram Höpken e Matthias Fuchs. Application of QR Codes in Online Travel Distribution. Em Ulrike Gretzel, Rob Law e Matthias Fuchs, editors, Information and Communication Technologies in Tourism 2010, páginas 137–148. Springer Vienna. Citado na pág. 2
- Ceipidor et al.(2009) Ugo B. Ceipidor, Carlo M. Medaglia, Amedeo Perrone, Maria De Marsico e Giorgia Di Romano. A museum mobile game for children using QR-codes. Em Proceedings of the 8th International Conference on Interaction Design and Children, IDC '09, páginas 282–283. Citado na pág. 2
- Chen e Wang(2009) Wen-Yuan Chen e Jing-Wein Wang. Nested image steganography scheme using QR-barcode technique. *Optical Engineering*, 48(5):057004. Citado na pág. 2
- Chu et al.(2007) Chung-Hua Chu, De-Nian Yang e Ming-Syan Chen. Image stablization for 2D barcode in handheld devices. Em Proceedings of the 15th International Conference on Multimedia, páginas 697–706. Citado na pág. 2, 8
- Coughlan et al. (2006) J. Coughlan, R. Manduchi e H. Shen. Cell phone-based wayfinding for the visually impaired. Em 1st International Workshop on Mobile Vision, in conjunction with ECCV, Graz, Austria. Citado na pág. 2, 8, 9

- **DensoWave()** DensoWave. QRcode.com. http://www.denso-wave.com/qrcode/index-e. html. Access on November 17, 2011. Citado na pág. 2
- Ebrahim et al. (2005) Y. Ebrahim, W. Abdelsalam, M. Ahmed e Siu-Cheung Chau. Proposing a hybrid tag-camera-based identification and navigation aid for the visually impaired. Em Second IEEE Consumer Communications and Networking Conference (CCNC), páginas 172 – 177. Citado na pág. 2, 8
- Felzenszwalb et al.(2010) Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester e Deva Ramanan. Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32:1627–1645. Citado na pág. 20
- **Freund e Schapire(1997)** Yoav Freund e Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of Computer and System Sciences*, 55(1):119 – 139. Citado na pág. 11, 16
- Liao e Lee(2010) Kuan-Chieh Liao e Wei-Hsun Lee. A Novel User Authentication Scheme Based on QR-Code. *Journal of Networks*, 5(8). Citado na pág. 2
- Lienhart e Maydt(2002) Rainer Lienhart e Jochen Maydt. An extended set of Haar-like features for rapid object detection. Em *IEEE ICIP*, páginas 900–903. Citado na pág. 12, 13, 27, 28
- Lienhart *et al.*(2002) Rainer Lienhart, Alexander Kuranov e Vadim Pisarevsky. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. MRL Technical Report, Microprocessor Research Lab, Intel Labs, Intel Corporation. Citado na pág. 30, 64
- Liu(2010) Shiyang Liu. Anti-counterfeit System Based on Mobile Phone QR Code and Fingerprint. Em Proceedings of the 2010 Second International Conference on Intelligent Human-Machine Systems and Cybernetics - Volume 02, páginas 236–240. Citado na pág. 2
- Liu et al.(2008a) Xu Liu, David Doermann e Huiping Li. A camera-based mobile data channel: capacity and analysis. Em Proceeding of the 16th ACM International Conference on Multimedia, páginas 359–368. Citado na pág. 8
- Liu et al.(2008b) Yue Liu, Ju Yang e Mingjun Liu. Recognition of QR Code with mobile phones. Em Chinese Control and Decision Conference (CCDC 2008), páginas 203–206. Citado na pág. 2
- Mohan et al.(2001) Anuj Mohan, Constantine Papageorgiou e Tomaso Poggio. Examplebased object detection in images by components. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23:349–361. Citado na pág. 19
- **Ohbuchi** et al.(2004) Eisaku Ohbuchi, Hiroshi Hanaizumi e Lim Ah Hock. Barcode readers using the camera device in mobile phones. Em *Proceedings of the 2004 International Conference on Cyberworlds*, páginas 260–265. IEEE Computer Society. Citado na pág. 2, 8
- **Ozcelik e Acarturk(2011)** Erol Ozcelik e Cengiz Acarturk. Reducing the spatial distance between printed and online information sources by means of mobile technology enhances learning: Using 2d barcodes. Computers & Education, 57(3):2077 2085. Citado na pág. 2
- Papageorgiou e Poggio(2000) C. Papageorgiou e T. Poggio. A trainable system for object detection. Intl. J. Computer Vision, 38(1):15â33. Citado na pág. 11

- Pârvu e Bălan(2009) Ovidiu Pârvu e Andrei G. Bălan. A method for fast detection and decoding of specific 2D barcodes. Em 17th Telecommunications Forum, TELFOR 2009, páginas 1137–1140. Citado na pág. 8
- Susono e Shimomura(2006) Hitoshi Susono e Tsutomu Shimomura. Using Mobile Phones and QR Codes for Formative Class Evaluation and Better Learning for College Students. Em Thomas Reeves e Shirley Yamashita, editors, Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2006, páginas 3026–3029. Citado na pág. 2
- Viola e Jones(2001) Paul A. Viola e Michael J. Jones. Rapid object detection using a boosted cascade of simple features. Em *IEEE Computer Society Conference on Computer* Vision and Pattern Recognition (CVPR), páginas 511–518. Citado na pág. 11, 12
- Walsh(2010) A. Walsh. QR codes using mobile phones to deliver library instruction and help at the point of need. *Journal of Information Literacy*, 4(1):55–64. Citado na pág. 2
- www.flickr.com(davidking) www.flickr.com, davidking. URL http://www.flickr.com/ people/davidking. Citado na pág. x, 54, 55
- www.flickr.com(gaku) www.flickr.com, gaku. URL http://www.flickr.com/people/gaku. Citado na pág. x, 57
- www.flickr.com(scott\_blake) www.flickr.com, scott\_blake. URL http://www.flickr.com/ people/scott\_blake. Citado na pág. x, 53
- www.flickr.com(cjsveningsson) www.flickr.com, cjsveningsson. URL http://www.flickr. com/people/cjsveningsson. Citado na pág. x, 55
- www.flickr.com(blech) www.flickr.com, blech. URL http://www.flickr.com/people/blech. Citado na pág. x, 52
- www.flickr.com(qriouscode) www.flickr.com, qriouscode. URL http://www.flickr.com/ people/qriouscode. Citado na pág. x, 56
- www.flickr.com(whoisstan) www.flickr.com, whoisstan. URL http://www.flickr.com/ people/whoisstan. Citado na pág. x, 58, 59, 60, 61
- Xue et al.(2010) Yinghua Xue, Guohui Tian, Rongkuan Li e Haitao Jiang. A new object search and recognition method based on artificial object mark in complex indoor environment. Em 8th World Congress on Intelligent Control and Automation (WCICA), páginas 6648–6653. Citado na pág. ix, 2, 9