

Person Re-Identification

Autor: Guilherme Zasyeki Machado

Sumário

- Introdução ao problema
- Visão geral ao processo de Re-Identificação
- Evolução estado da arte
- Datasets disponíveis
- Análise estado da arte
- Resultados
- Trabalhos Futuros
- Referências

Introdução ao problema (Motivação)

- Importante para *surveillance*
- Análise visual custosa, técnico precisa estar preparado
- Muita informação, várias câmeras para mão-de-obra limitada
- Importantes eventos/alvos podem passar despercebidos
- Falta de informações complementares. O que olhar? Contexto
- Visão computacional inicialmente entra em auxílio ao operador;

Introdução ao Problema (Dificuldade Encontradas)

- Dataset limitado
- Labels
- Generalizar
- Escalar
- Re-Identificação de longo prazo

Introdução ao problema (Desafio Computacional)

- Intra-class:
 - Iluminação
 - Posição ou Ângulo de Visão
 - Oclusão
- Inter-class:
 - Pessoas podem parecer iguais se vistas de ângulos diferentes
- Dois tópicos principais de pesquisa:
 - Feature Representation, deve ser discriminante para identidade, e sofrer pouco com variação de iluminação e ângulos de visão
 - Metric Learning, como aprender uma função de similaridade, ou, distância que seja robusta

Overview

Dataset podem ser de dois tipos:

- Single-shot: Apenas um par de imagens é comparado
- Multi-shot: Se a comparação for entre dois conjuntos de imagens

Overview



Fonte: [Karanam et al., 2016]

Evolução do Estado da Arte

Feature	Year	Metric	Year	Metric	Year
ELF [4]	ECCV 08	l_2		PCCA [8]	CVPR 12
LDFV [35]	ECCVW 12	FDA [45]	AE 1936	kPCCA [8]	CVPR 12
gBiCov [37]	BMVC 12	ITML [46]	ICML 07	LFDA [11]	CVPR 13
AlexNet-Finetune [40]	NIPS 12	MFA [47]	PAMI 07	SVMML [48]	CVPR 13
DenseColorSIFT [9]	CVPR 13	LMNN [49]	JMLR 08	kMFA [13]	ECCV 14
HistLBP [13]	ECCV 14	RankSVM [5]	BMVC 10	rPCCA [13]	ECCV 14
LOMO [18]	CVPR 15	PRDC [6]	CVPR 11	kLFDA [13]	ECCV 14
GOG [44]	CVPR 16	KISSME [7]	CVPR 12	XQDA [18]	CVPR 15

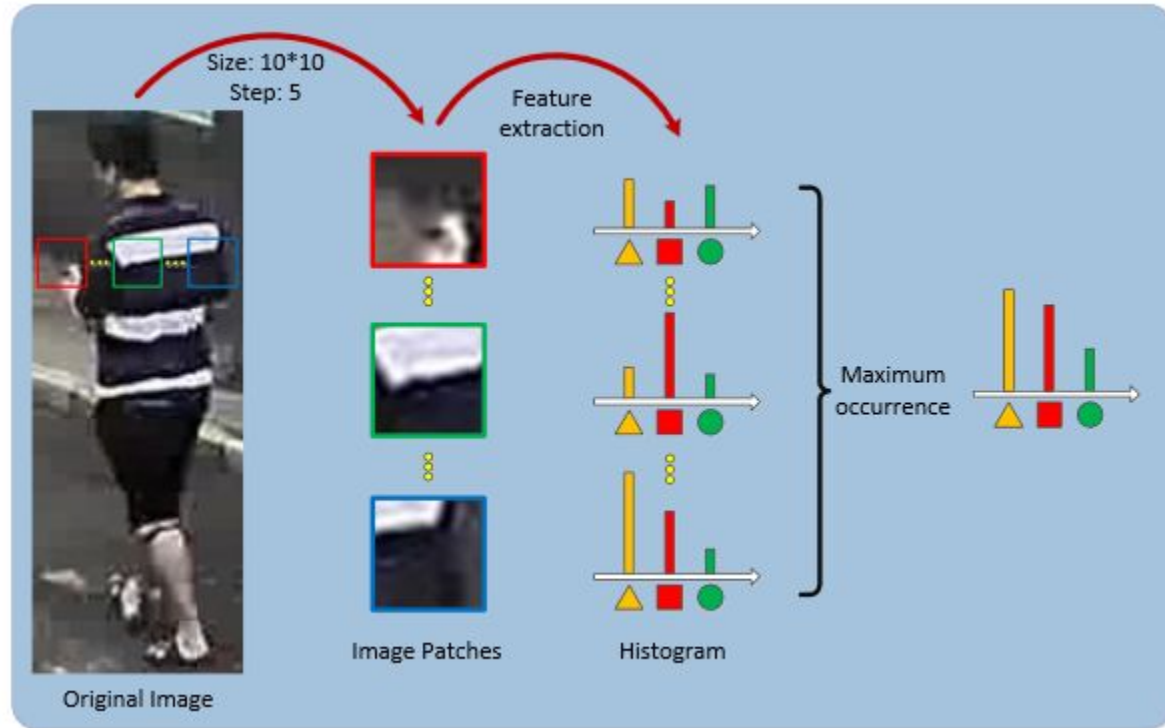
Fonte: [Karanam et al., 2016]

Principais Datasets

Dataset	# people	# BBox	# distractors	# cam	label	Attributes
VIPeR	632	1,264	0	2	hand	VV,IV
GRID	250	500	775	2	hand	VV,BC,OCC,RES
CAVIAR	72	1,220	0	2	hand	VV,RES
3DPeS	192	1,011	0	8	hand	VV,IV
PRID	178	24,541	0	2	hand	VV,IV
V47	47	752	0	2	hand	-
WARD	70	4,786	0	3	hand	IV
SAIVT-Softbio	152	64,472	0	8	hand	VV,IV,BC
CUHK03	1,360	13,164	0	2	DPM/hand	VV,DE,OCC
RAiD	43	6,920	0	4	hand	VV,IV
iLIDSVID	300	42,495	0	2	hand	VV,IV,BC,OCC
HDA+	74	2,976	0	12	ACF/hand	VV,IV,DE
Market1501	1,501	32,643	2,793+500K	6	DPM	VV,DE,RES
Airport	1,382	8,664	31,238	6	ACF	VV,IV,DE,BC,OCC

Fonte: [Karanam et al., 2016]

Feature LOMO (Local Maximal Occurrence)



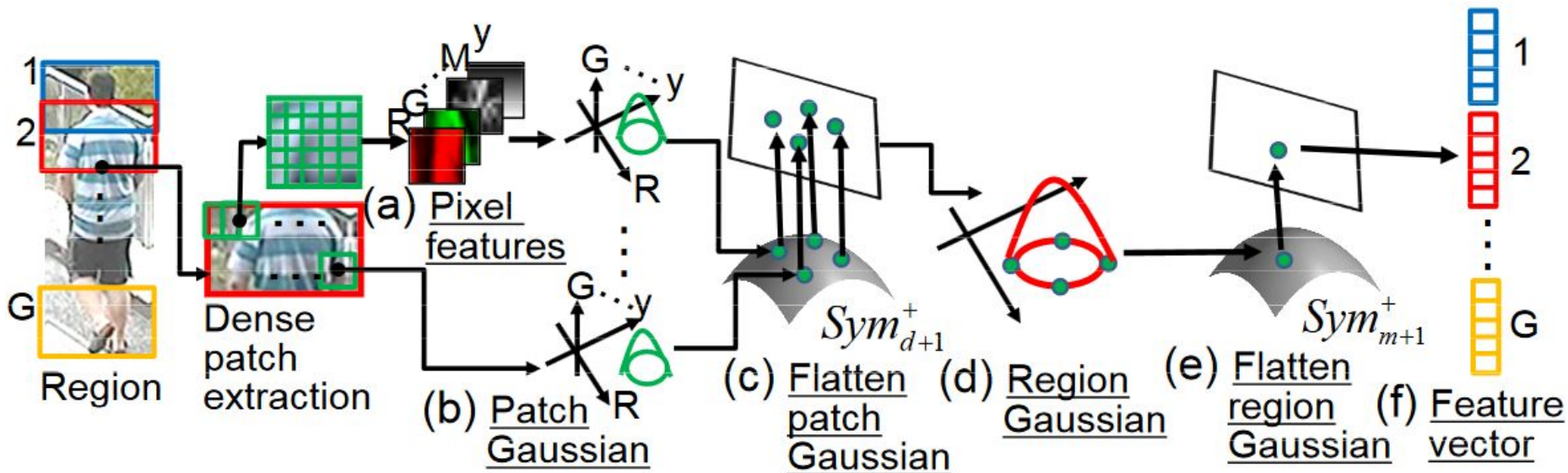
LOMO

- Pré Processa imagem com algoritmo Retinex multi escalado, para reduzir problema de iluminação.
- Aplica conversão de cores para HSV
- Adicionalmente aplica o algoritmo SILPT (Scale Invariant Local Ternary Pattern), uma variação do LBP
- O método constrói localmente o histograma dos pixels e pega o valor máximo na linha horizontal, para suprimir problemas de variação no ângulo de visão.

Feature GoG (Gaussian of Gaussian)

- Informações mais importantes na identificação:
 - Cor
 - Textura
- Descritor de Covariância (Lomo apenas a média)
 - Descreve a região de interesse como uma covariância de pixel features;
 - Forma natural, segundo o autor, de fusionar duas modalidades, cor e textura;
- Distribuição Gaussiana devido a importância da média local na discriminação de pessoas

Feature GoG (Gaussian of Gaussian)



Fonte: [Matsukawa et al., 2016]

Feature GOG (Pixel Feature)

- Extração de patches quadrados ($K \times K$)
- Para cada pixel extrai 8 features, normalizadas $[0,1]$

$$\mathbf{f}_i = [y, M_{0^\circ}, M_{90^\circ}, M_{180^\circ}, M_{270^\circ}, R, G, B]^T$$

Onde:

- y é a localização vertical no patch
- M são os gradientes de orientação multiplicados pela magnitude e quantizados nos bins $[0, 90, 180, 270]$
- R, G, B são os canais de cores da imagem

Feature GOG (Patch Gaussians)

- Para cada patch, sumariza-se os valores em uma distribuição Gaussiana, resultando em um vetor de features.

$$\mathcal{N}(\mathbf{f}; \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s) = \frac{\exp\left(-\frac{1}{2}(\mathbf{f} - \boldsymbol{\mu}_s)^T \boldsymbol{\Sigma}_s^{-1}(\mathbf{f} - \boldsymbol{\mu}_s)\right)}{(2\pi)^{d/2} |\boldsymbol{\Sigma}_s|}$$

- Onde $\boldsymbol{\mu}_s$, é o vetor médio, $\boldsymbol{\Sigma}_s$ é a matriz de covariância do patch s e $|\cdot|$ é o determinante da matriz.

Feature GOG (Flatten patch)

- Espaço Gaussiano não realiza operação Euclidiana (Riemannian manifold)
- Mas pode ser representado através de uma matriz SPD (Symmetric Positive Definite), também considerado um Rm. E então mapeado através da métrica log Euclidiana.
- O resultado é uma matriz simétrica, então, apenas o triângulo superior é armazenado.

Feature GOG (Region Gaussian)

- Patches locais sumarizados sem ordem, devido a variação na posição da pessoa
- Para suprimir os efeitos do background, dá-se peso maior para os patches mais centrais na imagem
- Calcula-se novamente a distribuição Gaussiana, agora dos patches, formando regiões
- Essa distribuição novamente deve ser representado através de uma matriz SPD e então mapeada para o espaço euclidiano
- Após obter todos as regiões gaussianas vetorizadas, concatena-se as G regiões, formando um descritor único

Feature GOG (Fusion)

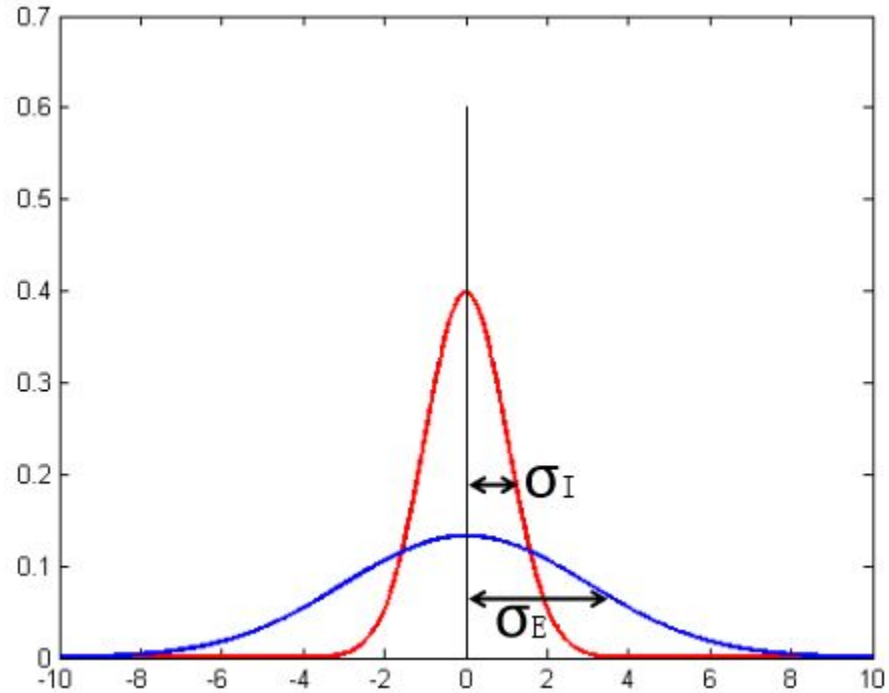
- Variação proposta pelos autores é mesclar diferentes representações de cor
- Foram usadas 4 representações {RGB, Lab, HSV, nRnG}, ao mesmo tempo
- Simplesmente concatenou-se as representações de cor, formando G_{Fusion} usado nos testes.
- Normalizou-se os descritores através da L2-norm, representado por:

$$\mathbf{z} = (\mathbf{z} - \bar{\mathbf{z}}) / \|\mathbf{z} - \bar{\mathbf{z}}\|_2$$

Matching XQDA

- XQDA (Cross-view Quadratic Discriminant Analysis)
- Baseado no problema de otimização de Fisher
 - Busca minimizar a dispersão intra-classe e maximizar a dispersão inter-classe
- Utiliza o parâmetro Ω_I , representando a variação intraclasse e Ω_E , representando a variação extraclasse, aplica distribuição Gaussiana
- Utiliza as informações de variância das classes ao invés de média

Matching XQDA



Resultados

- Normalmente avaliação realizada através de 2 métodos:
 - Rank-1: Percentual de imagens testadas e corretamente associadas ao seu grupo;
 - CMC Curve: São as chances de um match verdadeiro aparecer no top 1,2, ..., N do ranqueamento da lista, top 1 seria o equivalente a Rank-1.

Resultados

Datasets	Best Combination	1	5	10
VIPeR	GOG-XQDA	41.1	71.1	82.1
GRID	GOG-XQDA	21.5	38.1	49.4
3DPeS	GOG-kMFA _l	51.0	77.0	86.9
CUHK03	GOG-kLFDA _{exp}	62.1	88.7	94.2
HDA+	AlexNet-FDA	81.6	82.4	83.2
Market1501	GOG-kLFDA _l	58.6	79.4	85.7
Airport	GOG-XQDA	34.8	58.2	69.0
PRID	GOG-KISSME-SRID	91.5	97.8	98.8
V47	GOG-KISSME-SRID	100.0	100.0	100.0
CAVIAR	GOG-KISSME-RNP	55.6	79.6	95.6
WARD-12	GOG-KISSME-SRID	99.7	100.0	100.0
WARD-13	GOG-KISSME-SRID	96.0	98.6	99.1
SAIVT-38	GOG-KISSME-SRID	96.5	100.0	100.0
SAIVT-58	GOG-KISSME-RNP	72.6	89.9	93.0
RAiD-12	GOG-KISSME-SRID	100.0	100.0	100.0
RAiD-13	GOG-KISSME-SRID	91.9	94.8	96.2
RAiD-14	GOG-KISSME-SRID	95.7	96.2	99.5
iLIDSVID	GOG-KISSME-SRID	75.7	90.1	93.6

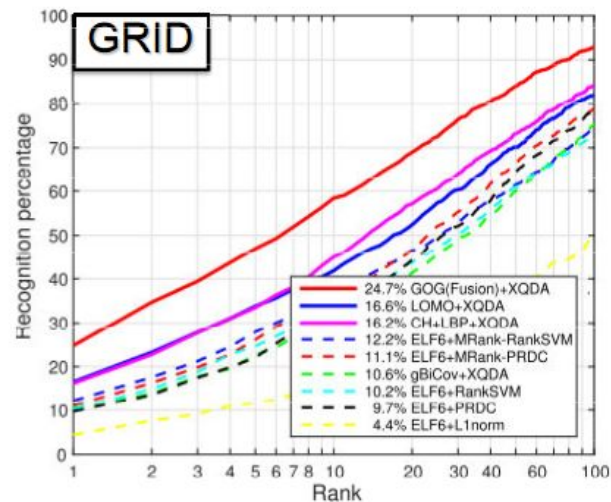
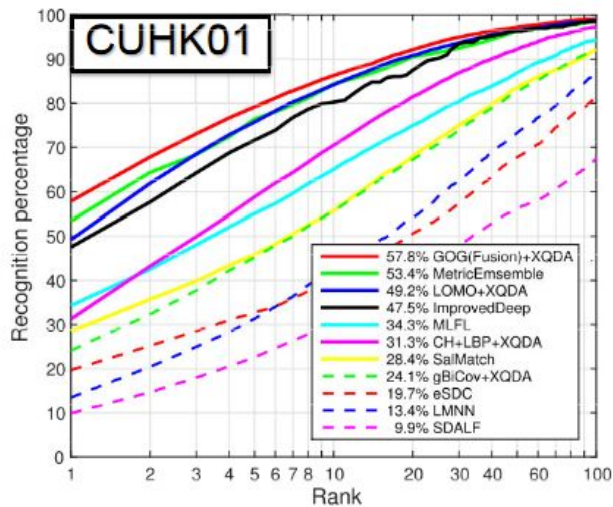
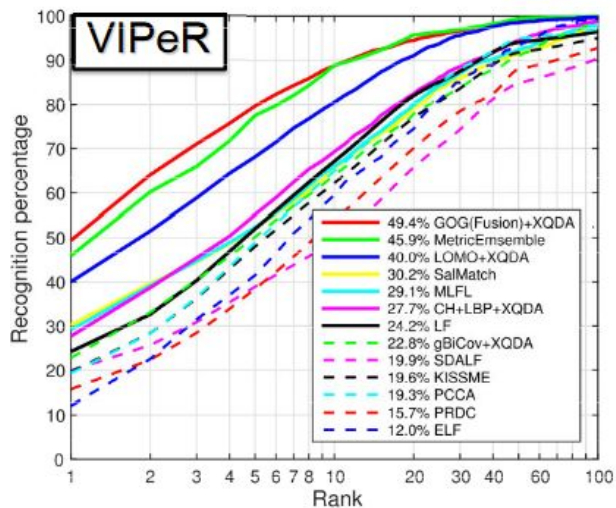
Fonte: [Karanam et al., 2016]

Resultados

Methods	Reference	VIPeR				CUHK01 (M=1)				CUHK01 (M=2)				PRID450S				GRID			
		r=1	r=5	r=10	r=20	r=1	r=5	r=10	r=20	r=1	r=5	r=10	r=20	r=1	r=5	r=10	r=20	r=1	r=5	r=10	r=20
GOG _{Fusion} +XQDA	Ours	49.7	79.7	88.7	94.5	57.8	79.1	86.2	92.1	67.3	86.9	91.8	95.9	68.4	88.8	94.5	97.8	24.7	47.0	58.4	69.0
MetricEnsemble	CVPR2015 [31]	45.9	77.5	88.9	95.8	53.4	76.4	84.4	90.5	-	-	-	-	-	-	-	-	-	-	-	-
LOMO+XQDA	CVPR2015 [23]	40.0	-	80.5	91.1	49.2	75.7	84.2	90.8	63.2	-	90.8	94.9	62.6	85.6	92.0	96.6	16.6	-	41.8	52.4
SCNCD	ECCV2014 [43]	37.8	68.5	81.2	90.4	-	-	-	-	-	-	-	-	41.6	68.9	79.4	87.8	-	-	-	-
Semantic	CVPR2015 [38]	31.1	68.6	82.8	94.9	32.7	51.2	64.4	76.3	-	-	-	-	43.1	70.5	78.2	86.2	-	-	-	-
SalMatch	ICCV2013 [45]	30.2	52	65	-	28.5	45	55	-	-	-	-	-	-	-	-	-	-	-	-	-
MLFL	CVPR2014 [47]	29.1	-	65.9	70.9	34.3	55	65	75	-	-	-	-	-	-	-	-	-	-	-	-

Fonte: [Matsukawa et al., 2016]

Resultados



Fonte: [Matsukawa et al., 2016]

Trabalhos Futuros

- Deep hierarchy Gaussian
- Testar métodos de conjuntos (ensemble methods) com os descritores GOG
- Para o contexto multi-shot, desenvolver métodos que usem a distribuição temporal podem trazer resultados
- Desenvolver dataset mais complexos e completos para uso de CNN

Referências

[Karanam et al., 2016] Karanam, S., Gou, M., Wu, Z., Rates-Borras, A., Camps, O. I., and Radke, R. J. (2016). A comprehensive evaluation and benchmark for person re-identification: Features, metrics, and datasets. CoRR, abs/1605.09653.

[Matsukawa et al., 2016] Matsukawa, T., Okabe, T., Suzuki, E., and Sato, Y. (2016). Hierarchical gaussian descriptor for person re-identification. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

Referências

[Liao et al., 2015] Liao, S., Hu, Y., Zhu, X., and Li, S. Z. (2015). Person re-identification by local maximal occurrence representation and metric learning. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2197–2206

[Gong et al., 2014] Gong, S., Cristani, M., Yan, S., and Loy, C. (2014). Person Re-Identification. Advances in Computer Vision and Pattern Recognition. Springer London.