

Relatório de Exercício Programa 1

Busca de Imagens

Rafael de Oliveira Lopes Gonçalves {*rafaellg@vision.ime.usp.br*}

Rodrigo Suzuki Okada {*rsuzuki.okada@gmail.com*}

William Mizuta {*william.mizuta@gmail.com*}

Setembro 2010

1 Introdução

O objetivo deste trabalho é, através do OpenCV, criar e avaliar um método de busca de imagens pela sua similaridade, bem como comparar o método desenvolvido com um método de referência. No caso, o algoritmo, dado uma entrada, deve retornar um conjunto de imagens que *ache parecido* com a entrada.

Note que o conceito de similaridade em imagens não é trivial. A proposta inicial do algoritmo nada diz a respeito sobre como quantificar o quanto uma imagem difere de outra, ou então, quais características das imagens as fazem torná-las mais similares. Assim, para este trabalho, técnicas para analisar certas características das imagens foram testadas até encontrar uma que tivesse resultados satisfatórios.

2 Metodologia

2.1 Algoritmos

Para a análise do novo método, foi necessário estabelecer um critério de comparação com um outro algoritmo, bem como definir qual este seria. No caso, foi pré-definido o algoritmo base como sendo a intersecção de histogramas das imagens. Mais especificamente, foi definido que a imagem, convertida para o espaço de cor *HSV*, teria seus canais *H* e *S* discretizados em 64 níveis para a montagem do histograma para que, então, possa ser feito um cálculo de similaridade pela intersecção.

Note que o método descrito analisa uma característica global da imagem, prontamente disponível, sem analisar o conteúdo semântico ou posicionamento dos objetos. Assim, para comparação, foram levantadas algumas alternativas que avaliassem tais características.

2.1.1 Método de Referência

O método de referência é utilizar a intersecção de histogramas entre duas imagens. Mais especificamente, as imagens são convertidas para o espaço de cor *HSV*, e então, os canais *H* e *V* são quantizados em 64 níveis cada, para então construir os histogramas que serão comparados via intersecção.

2.1.2 SURF

Uma tentativa inicial foi a utilização do SURF como método para analisar as imagens através de *features* (ou então, pontos de interesse) invariantes com escala e rotação, para então, casar os pontos de duas imagens. Fotos de um mesmo objeto que contenham pontos de vistas próximos devem gerar imagens similares, e assim, podem ser casadas.

Porém, este algoritmo mostrou-se de difícil avaliação. Primeiramente, precisou-se de um método para quantificar o nível de similaridade entre duas imagens para que uma ordenação seja possível. Pensou-se em utilizar a quantidade de pontos casados para quantificar a similaridade, mas o como as fotos do banco e as de entrada tem pontos de vistas com grande diferença acaba aumentando os falsos positivos nos resultados.

Além disso, não se chegou a um consenso de como avaliar a qualidade dos resultados para que a ordenação seja possível. Pensou-se em um modelo em que o nível de similaridade era proporcional à quantidade de pontos encontrados e que foram casados, apesar de tal heurística não ter se mostrado eficiente.

2.1.3 Análise por Região

Outra método sugerido foi analisar as imagens em partes, aonde cada imagem é dividida em $N \times N$ blocos, numerados de 0 a $N \times N - 1$, aonde o bloco i é aquele presente na linha $i \div N$ e coluna $i \% N$. A comparação entre duas imagens, então, é feita comparando seus blocos de mesma posição, extraíndo um valor numérico de cada um quantificando a diferença entre blocos. A diferença das imagens, então, torna-se as médias das diferenças entre cada bloco.

Assim:

$$dif.nn(a, b) = \frac{\sum_{i=0}^{N \times N - 1} dif(bloco(a, i), bloco(b, i))}{N \times N}$$

Aonde $dif(i, j)$ é uma função que retorna a diferença entre duas imagens i e j , e $bloco(x, i)$, uma função que extrai o bloco i da imagem x . No caso, a função $dif(i, j)$ é a intersecção de histogramas utilizada no método de referência, porém, utilizando 32 níveis ao invés de 64. Para este experimento, foi utilizado $N = 2$.

Porém, o centro da imagem, normalmente, possui grande relevância na sua identificação. O método dos $N \times N$ blocos não contempla este tipo de informação. Assim, além destes blocos, foi analisando um bloco com a metade da altura e largura, posicionada ao centro da imagem. Este bloco tem peso 0.4 em cima da diferença de todos os outros blocos.

$$dif.img = \frac{dif.nn(a, b) + 0.4 \times dif(centro(a), centro(b))}{1.0 + 0.4}$$

O resultado de todas as funções dif retornam valor de 0 a 1, aonde 1 significa totalmente diferentes, e 0, totalmente similares.

2.2 Avaliação

A base de imagens que seria pesquisada continha 120 imagens pré-definidas, todas de tamanho idêntico, em formato JPG. Todas também continham informações de cor.

Para avaliar o resultado de cada método, foi utilizado um *Ground-Truth* como referencial. Este oráculo diz qual seria as imagens mais parecidas para uma imagem qualquer. Neste trabalho, foram utilizadas 10 imagens de entrada pré-definidas, cujo oráculo foi gerado a partir de opiniões de alunos de MAC5915, aonde era pedido que cada um listasse as 10 imagens mais parecidas com a imagem de entrada. As imagens mais votadas tornaram-se o Ground-Truth. No geral, oráculo foi formado por imagens cujo valor semântico melhor se aproximava da foto de entrada, isto é, tinham objetos que podem receber um mesmo nome.

3 Resultados

Na tabela 1, pode-se verificar a saída do algoritmo utilizado como base de comparação, verificando os 10 melhores resultados para cada imagem de entrada.















































































































Entrada	1	2	3	4	5	6	7	8	9	10
										
										
										
										
										
										
										
										
										
										

Tabela 1: Resultado do algoritmo de referência

Na tabela 2, o resultado obtido pelo novo método, utilizando parâmetros de localidade e auto-nivelamento.






































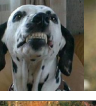



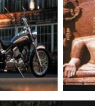

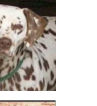



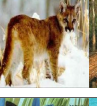

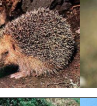
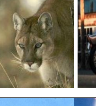

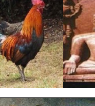




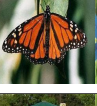
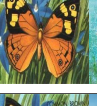

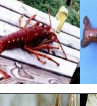
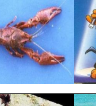



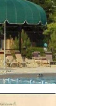



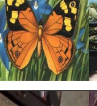


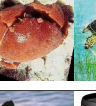

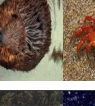

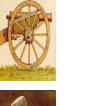




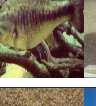



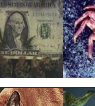



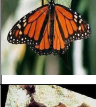

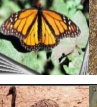


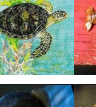
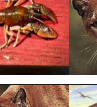
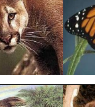
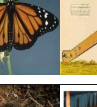
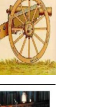



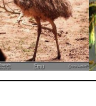


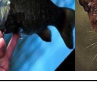


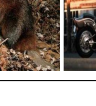

Entrada	1	2	3	4	5	6	7	8	9	10
										
										
										
										
										
										
										
										
										
										

Tabela 2: Resultado do algoritmo proposto

4 Análise de Resultados

O projeto propõe-se a solucionar dois problemas da área de busca de imagens: encontrar as n imagens mais similares, que será explicado na seção 4.1; e encontrar um classificador para identificar as imagens que pertencem ao mesmo grupo da imagem de referência, explicado na seção 4.2.

4.1 Busca das mais similares

4.1.1 Ground truth

Realizou-se uma entrevista na qual cada um teve que eleger, em ordem crescente, as dez imagens do banco de imagens que eles achavam mais parecidas na opinião de cada um. Com isso, tem-se uma referência que servirá como referência para os algoritmos implementados. O resultado da pesquisa pode ser visto na tabela 3:

















































































































Entrada	1	2	3	4	5	6	7	8	9	10
										
										
										
										
										
										
										
										
										
										
										

Tabela 3: Ground truth

4.1.2 Medida

Para verificar a eficácia dos algoritmos em relação ao *ground truth*, criou-se a seguinte medida:

$$T = \sum_{i=1}^{10} (10 - |l_i - p_i|) * (11 - p_i) \quad (1)$$

onde l_i é a posição da imagem i no resultado do algoritmo e p_i é a posição da mesma imagem no *ground truth*. Caso a imagem não esteja no *ground truth*, o valor de p_i é 11. Essa métrica atribui pesos maiores para imagens que são mais parecidas ($11 - p_i$) e verifica a distância entre a posição da imagem no resultado do algoritmo e no *ground truth* ($10 - |l_i - p_i|$). O resultado pode ser visto na tabela 4.









Entrada	Histograma Global	Momentos em Janelas	Histograma em Janelas	Surf
	0.016364	0.036364	0.336364	0.105455
	0.220000	0.116364	0.356364	0.101818
	0.036364	0.300000	0.492727	0.000000
	0.000000	0.278182	0.081818	0.000000
	0.032727	0.178182	0.000000	0.147273
	0.156364	0.210909	0.214545	0.000000
	0.116364	0.029091	0.140000	0.007273
	0.000000	0.114545	0.130909	0.081818
	0.490909	0.278182	0.529091	0.116364
	0.196364	0.000000	0.087273	0.192727
Média	0.126545	0.154182	0.236909	0.075273
Desvio Padrão	0.144383	0.105883	0.172821	0.066294

Tabela 4: Comparando a medida de similaridade do algoritmos estudados

4.2 Classificador de busca

5 Análise de Resultados

Primeiramente, percebe-se que nenhum dos métodos propostos chegaram a concordar com as imagens do oráculo, apresentando uma taxa de falsos positivos elevada. Este fato evidencia diferenças entre as maneiras como um ser humano interpreta uma imagem e os métodos propostos, que analisam valores extraídos diretamente da imagem, e não sua semântica, que pode ser notada pelo *ground truth* escolhido, aonde os itens mais votados são aqueles que representam um mesmo tipo de objeto da imagem de entrada, independentemente do ângulo de visão, condições de luminosidade, posição ou escala do objeto.

O método de momentos de janelas, quando comparado ao método de referência, mostrou-se mais estável, com taxa maior de acertos com menor flutuação. Ainda assim, foi inferior à taxa de acertos do método de histogramas por janelas, que, contudo, apresentou o maior desvio entre os métodos avaliados. Aliás, as janelas mostraram uma melhoria com relação a uma análise global, até por utilizar características regionais

de um local para analisar a forma como as cores estão dispostas, mesmo que de maneira simples.

O SURF teve o pior desempenho dentre os novos métodos, com taxa de acertos menores do que o método original, e com um desvio valor próximo à própria média. Não se esperou uma taxa de acerto altas, até pelo banco de imagens, compostos por objetos sob pontos de vistas e locais muito diferentes. O SURF, apesar de útil na junção de imagens panorâmicas, aonde cada imagem contém uma parte da imagem total, não é robusta para este tipo de operação de busca.

Vale notar que o método original tem desvio superior à própria média, o único método que teve tal característica.

Por fim, nota-se que a taxa de acertos por imagem teve grande variância. Enquanto a imagem de borboleta teve acertos muitos superiores à média, as imagens da formiga, jacaré e peixe tiveram acertos, no geral, bem abaixo da média.

6 Conclusão

Este projeto, no que se refere ao aprendizado do OpenCV, foi de grande valia, aonde múltiplos métodos foram desenvolvidos através de funcionalidades básicas fornecidas, permitindo o uso da criatividade para criar formas com que as imagens deveriam ser buscadas. Apesar de não usar algoritmos complexos, excessão feita ao Surf, o aprendizado da manipulação de canais de cores, valores de pixels entendimento da estrutura básica do OpenCV foi o maior benefício para os autores.

A análise dos resultados também teve sua contribuição, aonde foi visto uma análise através de um oráculo que pode ser definido como verdade absoluta do sistema ideal. Definir um parâmetro de comparação no início foi um passo importante, visto que é perfeitamente possível criar modelos de comparação após criar um novo algoritmo, cujo resultado é parcialmente forjado por uma análise tendenciosa.

Por fim, vale notar as taxas encontradas, que foram bem abaxias do que pode ser considerado ideal. Apesar de não ser o objetivo principal do projeto, tal resultado exemplifica as dificuldades da visão computacional em desenvolver métodos precisos e computacionalmente eficientes para executar suas tarefas.