

# Quantização de imagens para geração de descritores compactos de cor e textura

Luciana C. Escobar, Moacir P. Ponti Jr.  
Instituto de Ciências Matemáticas e de Computação  
Universidade de São Paulo  
13560-970 São Carlos, SP, Brasil  
Email: moacir@icmc.usp.br, lcepej@grad.icmc.usp.br  
Web page: <http://www.icmc.usp.br/~moacir>

**Resumo**—A classificação de imagens envolve diversas etapas, como aquisição, pré-processamento, extração de características e classificação. O bom desempenho do classificador e a sua velocidade de processamento dependem das características extraídas. Os parâmetros dos passos anteriores também podem influenciar significativamente os resultados. Nesse estudo foi estudada a relação entre o nível de quantização de imagens e os resultados da classificação. Três descritores visuais foram obtidos a partir de imagens com 8, 16, 32, 64, 128 e 256 cores. Na literatura, a escolha do número de cores é feita comumente de forma arbitrária. Por isso investigamos como os descritores e textura e cor se comportam usando imagens com diferentes números de cores. Demonstramos que, para descritores de cor, a complexidade de tempo e espaço aumenta mais rápido do que a acurácia conforme o número de cores aumenta. Os resultados apresentados podem beneficiar diversas aplicações, oferecendo uma forma de escolher métodos rápidos e compactos para extração de atributos.

**Keywords**—Descritores visuais, Quantização, Reconhecimento de Padrões.

**Abstract**—Image classification comprises several steps such as acquisition, preprocessing, feature extraction and then classification. In this task the extracted features plays a significant role on the accuracy and speed of the system. The parameters of the steps before the extraction can also change significantly the results. In this paper we investigated the relation between the quantization level of images and the classification results. Three descriptors were obtained from images quantized with 8, 16, 32, 64, 128 and 256 colors. While the choice of the number of colors is often arbitrary in the literature, we studied how color and texture descriptors behave by using images with different number of colors. It is shown that, for color descriptors, the time and space increase faster than the accuracy when the number of color increase. The results can benefit several applications by offering a way to choose compact and fast methods for feature extraction.

**Keywords**—Visual descriptors, Quantization, Pattern Recognition.

## I. INTRODUÇÃO

A análise e o reconhecimento de padrões em imagens são importantes para aplicações variadas em visão computacional e recuperação de imagens baseada em conteúdo. Imagens naturais são casos específicos de difícil classificação automática, por causa de suas características (estatísticas e estruturais) que dificultam seu processamento e análise. Por exemplo, objetos e cenas naturais de uma mesma categoria podem ter estrutura e estatística bastante diversas devido às diferentes condições

de aquisição. Contudo, há muitas aplicações realizadas sobre imagens naturais, o que torna importante o estudo e a avaliação de métodos capazes de lidar com essas dificuldades.

A classificação de imagens possui diversas etapas. A etapa inicial é a aquisição, onde a imagem é capturada por algum dispositivo de imagem ou vídeo. Na aquisição são definidos os parâmetros de amostragem, que definem a resolução da imagem digital, e quantização, que define a sua profundidade (em bits por pixel). Quando a imagem é utilizada como entrada de um sistema de reconhecimento de padrões esses parâmetros podem ser bastante diferentes de quando as imagens são adquiridas para visualização. Esses parâmetros podem influenciar na acurácia e velocidade e portanto devem ser estudados para cada aplicação [1].

Após a aquisição, uma etapa de pré-processamento busca melhorar a imagem antes de ela ser processada, com tratamento de ruídos, realce, segmentação, entre outros métodos; o pré-processamento não é obrigatório mas pode melhorar consideravelmente o resultado final [2].

A seguir, é necessário escolher uma metodologia que permita comparar diferentes imagens por alguma medida de similaridade ou dissimilaridade. Dependendo da metodologia utilizada, este passo é chamado de extração de características ou de descrição. Formalmente, um descritor é definido como um par  $(E; D)$ , onde  $E$  é o algoritmo para realizar a extração do vetor de características e  $D$  é uma função responsável pela comparação de dois vetores.

Na literatura, existem vários descritores utilizados em aplicações diversas. Cada descritor tenta capturar certas características presentes nas imagens como: cor, textura, forma, entre outros. No reconhecimento de objetos em imagens naturais, o estado da arte utiliza informações de cor, gradiente, textura e sua localização espacial num framework complexo chamado *bag-of-visual-words* [3]. No entanto, há uma grande variedade de descritores visuais, em especial de textura e cor disponíveis e que são simples e suficientes para diversas aplicações [4].

Em algumas aplicações, como sistemas embarcados e computação móvel, é importante utilizar descritores compactos e rápidos de computar, com informações relevantes na menor quantidade de dados possível. Essas características são fundamentais para permitir processamento local e em tempo real, e viabilizar a transmissão via rede sem fio com poucos pacotes.

Essa necessidade vem de encontro a aplicações recentes como reconhecimento facial [5] e de reconhecimento de códigos de barras [6]. Apesar da crescente melhoria do hardware desses sistemas, diversas limitações ainda permanecem, como a instabilidade e velocidade da rede de dados e o consumo de energia e, portanto a investigação de métodos rápidos com estruturas de dados compactas é um tema relevante e atual [7].

Em um estudo anterior foi demonstrado que o uso de imagens com resolução reduzida pode beneficiar aplicações com grande número de imagens, permitindo processar e extrair características de forma mais rápida sem perda na acurácia da classificação [1].

Neste artigo são utilizados dois descritores de cor e um descritor de textura e é analisada a acurácia da classificação em imagens com diferentes níveis de quantização (quantidade de cores: 8, 16, 32, 64, 128, 256). O resultado deste estudo mostra até onde é possível diminuir a quantidade de cores, para obter descritores compactos e classificação mais rápida, sem perder informações relevantes para a classificação.

O restante do artigo está organizado da seguinte forma:

- Seção II: revisão da literatura incluindo os descritores visuais utilizados e o classificador de padrões
- Seção III: metodologia utilizada na investigação, bases de imagens e descrição dos experimentos.
- Seção IV: apresentação dos resultados e discussão.
- Seção V: comentários finais sobre os resultados, conclusões e trabalhos futuros.

## II. REVISÃO DA LITERATURA

Nesta seção serão apresentados conceitos importantes para o entendimento deste estudo.

### A. Descritores visuais

*GCH (Global Color Histogram)*: representa a distribuição global de cores na imagem, ou seja, representa a probabilidade de um pixel ser de uma determinada cor. O primeiro passo para obter essa frequência é quantizar a imagem em uma determinada quantidade de cores; após isso, o histograma é calculado com base na cor de cada pixel. Este método é simples, mas bastante útil em diversas aplicações [8].

*BIC (Border/Interior Classification)*: método de extração de características de cor que consiste em classificar cada pixel da imagem como borda ou interior; um pixel é considerado interior se seus vizinhos (em geral 4-conectados) possuem a mesma cor que o pixel em questão, caso contrário o pixel é considerado como borda [9].

Antes da análise dos pixels a imagem deve ser requantizada para uma menor quantidade de cores (o artigo original utiliza 64). Este descritor é composto por dois histogramas, cada um de tamanho igual à quantidade de cores utilizada na quantização, sendo um histograma para borda e outro para interior. Através da análise da quantidade de pixels de uma determinada cor nos dois histogramas podemos chegar às seguintes conclusões:

- Se a quantidade de pixels borda é menor do que a quantidade de pixels interior, então a cor é distribuída em amplas regiões com formato regular;
- Se a quantidade de pixels interior é menor do que a quantidade de pixels borda, então pelo menos uma destas afirmações é verdadeira:
  - 1) A cor é distribuída em regiões pequenas e conectadas pela borda;
  - 2) A cor é distribuída em regiões grandes com formato irregular;
  - 3) A cor pertence a uma região com predominância de textura.

*Descritores de Haralick*: Haralick et al. (1973) [10] descrevem um método de classificação baseado em texturas realizado a partir de cálculos estatísticos de segunda ordem (relação entre dois pixels: o de referência e vizinhos), que definem diversas características obtidas através de uma matriz de coocorrência; a matriz de coocorrência possui os valores da ocorrência de cores em uma determinada imagem. Essas características são utilizadas para diferenciar texturas que não seguem um determinado padrão de repetição. Nesse estudo foram extraídos os seguintes descritores:

- Máxima probabilidade: resposta mais forte da imagem;
- Energia: descritor da uniformidade;
- Homogeneidade: descritor da uniformidade da distribuição de pixels;
- Contraste: variação local das cores entre pixels;
- Entropia: descritor da aleatoriedade das cores em uma imagem;
- Correlação: descritor da correlação de cores entre pixels vizinhos.

### B. Classificador OPF

Uma vez concluída a etapa de extração de características, segue a etapa da classificação. Para isso é necessário um método de treinamento que aprende com os dados previamente rotulados a fim de poder classificar dados desconhecidos de novas imagens.

Os rótulos atribuídos aos dados são chamados de classes; uma classe possui objetos similares entre si e, logicamente, cada classe é distinta das outras. Em alguns casos a distinção entre as classe é bastante nítida, contudo existem classes em que há uma certa confusão entre os elementos, devido ao fato dos vetores de características de classes diferentes ocuparem lugares próximos no espaço de características. Por isso é preciso escolher um classificador que seja robusto o suficiente para lidar com diferentes situações.

Neste estudo foi utilizado o classificador OPF (Optimum-Path Forest) [11]. O OPF calcula uma floresta de caminhos ótimos modelada como um grafo gerado a partir do espaço de atributos; essa floresta reduz o problema de classificação de padrões ao de encontrar a árvore que minimiza o custo até a sua raiz (chamada de protótipo). As fronteiras de uma classe são formadas pelos elementos mais representativos de cada classe (os protótipos), elementos esses do conjunto de

treinamento. Para um novo objeto, o classificador calcula o caminho de menor custo de cada nó da árvore até a raiz e classifica esse novo objeto com a mesma classe do nó com o caminho de menor custo.

O valor de um caminho é calculado por uma função dependente da aplicação utilizada e é realizado com base na distância entre um objeto e o seu vetor descritor. Após o término desse processo, várias árvores de caminho são geradas, e a união dessas árvores é denominada floresta de caminhos ótimos.

O OPF foi escolhido por possuir algumas vantagens como:

- Ter demonstrado ser mais rápido e possuir eficácia similar à classificadores complexos como SVM (Suport Vector Machines), Bayesianos e Redes Neurais;
- Suporta problemas multiclasse de forma nativa;
- Possibilita o uso de qualquer função de distância;
- Não necessita de ajustes em parâmetros;
- Suporta algum grau de sobreposição entre as classes.

O fato de não possuir parâmetros é ideal para nosso estudo pois não é necessário se preocupar com a otimização do classificador. Além disso as diversas aplicações que o utilizaram apresentaram bom desempenho [12], [13]. Para maiores detalhes desse classificador ver o trabalho de Papa et al. [11].

### III. METODOLOGIA E EXPERIMENTOS

Esta seção contém detalhes das bases de imagens utilizadas, características da implementação e dos experimentos.

#### A. Bases de imagens

Com objetivo de investigar bases de imagens naturais com características diversas, obtidas sob diferentes condições e com muitas classes, foram selecionados os seguintes conjuntos:

- **Corel:** 1000 imagens, 10 classes sendo elas: África, cavalos, comidas, construções, dinossauros, elefantes, flores, ônibus, montanhas e praia;
- **Covers:** 99 imagens, 33 classes de capas de CD variados;
- **Paintings:** 99 imagens, 33 classes de obras de arte variadas obtidas no European Gallery of the Cantor Arts Center.

A base de imagens Corel é um subconjunto de imagens gerado por [14] e disponível na web <sup>1</sup>, com resolução 384×256.

As bases Covers e Paintings foram construídas com vistas a aplicações em dispositivos móveis, com imagens obtidas por celulares. Covers possui imagens com resolução 1280×960 e Paintings com resolução 2048×1536. As imagens de teste dessas bases foram obtidas em condições de baixa iluminação, com rotação, oclusão parcial, ruído, desfocamento, e, no caso da base Covers, utilizando flash do tipo LED <sup>2</sup>.

Para estudar e avaliar o impacto de diferentes níveis de quantização no desempenho da classificação as imagens foram quantizadas com as seguintes quantidades de cores: 8, 16, 32, 64, 128 e 256.

Exemplos de imagens das bases utilizadas estão na Figura 1.

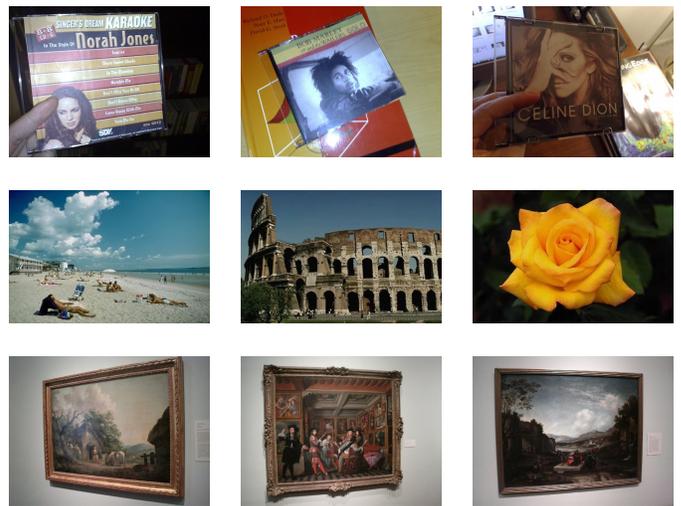


Figura 1. Exemplos de imagens de três classes diferentes para cada base de imagens. Covers: primeira linha, Corel: segunda linha, Paintings: terceira linha

#### B. Implementação

A biblioteca OpenCV [15] foi utilizada em conjunto com a linguagem C para realizar a quantização das imagens e a extração das características. A quantização foi realizada por transformações em cada pixel da imagem através de operações nos seus bits de cada canal de cor RGB a fim de obter um único canal de cor que possuísse informações dos três canais de cor originais (R, G e B), conforme descrito em [1].

Em relação ao descritor BIC, todos os pixels da borda externa da imagem foram considerados pixel borda, e considerados para classificação borda/interior os 4 vizinhos de cada pixel. Os vetores de características GCH e BIC tem tamanho diferente a depender da quantização utilizada (quanto menos cores, menor o vetor).

Para os descritores de Haralick a matriz de coocorrência foi computada verificando os pixels com distância 1 e foram utilizadas as 6 características descritas na seção II-A. Por isso o vetor de características possui 6 atributos para qualquer nível de quantização.

Todos os vetores de características foram normalizados. A distância euclidiana foi utilizada em conjunto com os extratores Haralick e GCH e a distância logaritmica [9] com o extrator BIC.

#### C. Experimentos

A fim de obter dados para analisar o impacto da quantidade de cores na classificação de imagens, os experimentos foram repetidos 20 vezes e realizados com a seguinte configuração:

- **Corel:** 30 instâncias por classe para treinamento e 70 para teste;
- **Covers e Paintings:** 2 instâncias por classe para treinamento e 1 para teste;

### IV. RESULTADOS E DISCUSSÃO

Os resultados obtidos neste estudo estão representados em três gráficos, um para cada base de imagem contendo a

<sup>1</sup><http://wang.ist.psu.edu/docs/related/>

<sup>2</sup><http://www.stanford.edu/class/ee368/>

acurácia da classificação obtida pelos conjuntos de atributos de cor e textura usando imagens em diferentes quantizações. Os resultados observados para as bases de imagens Covers, Corel e Paintings podem ser vistos nas Figuras 2, 3 e 4, respectivamente. Os tempos médios para extração de características de cada imagem por cada método são apresentados na Figura 5.

Os resultados da acurácia da classificação foram relativamente diferentes para cada base de imagem, devido ao: número de amostras disponíveis por classe, número de classes, resolução, condições de aquisição, etc.

A base de imagens Covers foi a que apresentou os piores resultados, que variam entre 48% e 53%. A base é desafiadora pela grande quantidade de classes (33) e baixo número disponível de amostras por classe para treinamento (apenas 2). Os melhores resultados foram obtidos para os três descritores quando a imagem possuía 8 cores; isto acontece provavelmente porque quando a imagem é quantizada para 8 cores, as interferências como fundo e iluminação sofrem alterações em suas cores e deixam de serem tão marcantes. Por isso não conseguimos obter conclusões a partir dessa base.

Nas bases Corel e Paintings, observamos comportamentos similares, com resultados entre 50 a 85% para a base Paintings e entre 60% e 80% para a base Corel. A acurácia da classificação quando utilizados os descritores BIC e GCH cresce conforme a quantidade de cores aumenta. Tal comportamento era esperado, pois o BIC e o GCH são descritores de cor e, maior número de cores significa mais informações para esses descritores utilizarem. Com 8 cores o resultado é relativamente menor se comparado com os demais, porém com 16 cores o resultado está em um nível aceitável para os dois descritores (aproximadamente 75%). É interessante observar que para os descritores de cor:

- O tempo para extrair os atributos e o tamanho do vetor crescem de forma quadrática conforme aumenta o número de cores.
- A acurácia da classificação cresce de forma lenta a partir de 16 cores.

Nestas duas bases de imagens, o descritor Haralick possui um melhor desempenho quando a quantidade de cores é menor. O descritor é baseado em texturas e com muitas cores pode existir certa confusão entre quais informações de cor são relacionadas à textura e quais não são. A acurácia varia de 50% a 70% para a base Paintings e de 60% a 65% para a base Corel. Por extrair quantidade de atributos fixa para qualquer nível de quantização, gasta um tempo médio em cada base, sem aumentos ou diminuições significativos.

## V. CONCLUSÕES

Com o objetivo de estudar a viabilidade do uso de descritores compactos para o reconhecimento de padrões visuais foram realizados experimentos para avaliar o uso de diferentes parâmetros de quantização na classificação de imagens naturais. A utilização dos descritores BIC, GCH e Haralick se justificou pelo fato de serem rápidos durante seu processamento e de serem compactos para quantidade de cores menores (o Haralick é compacto em qualquer nível de quantização).

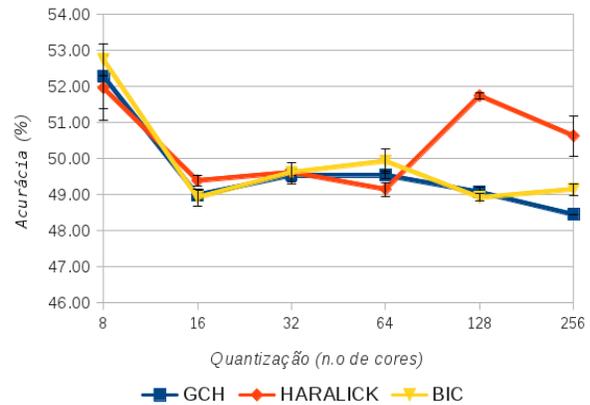


Figura 2. Resultados utilizando a base COVERS para os três descritores e diferentes quantizações

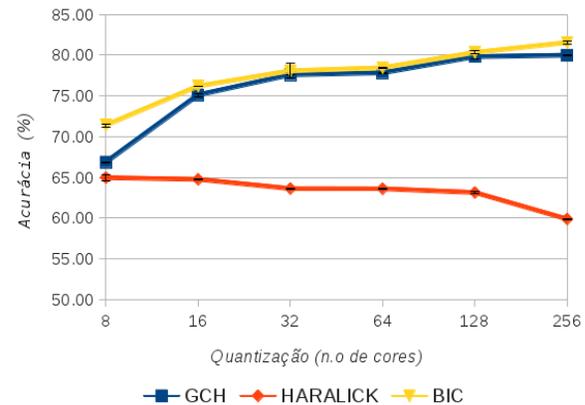


Figura 3. Resultados utilizando a base COREL para os três descritores e diferentes quantizações

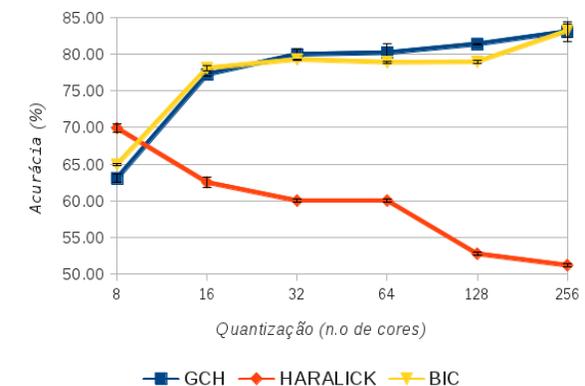


Figura 4. Resultados utilizando a base PAINTINGS para os três descritores e diferentes quantizações

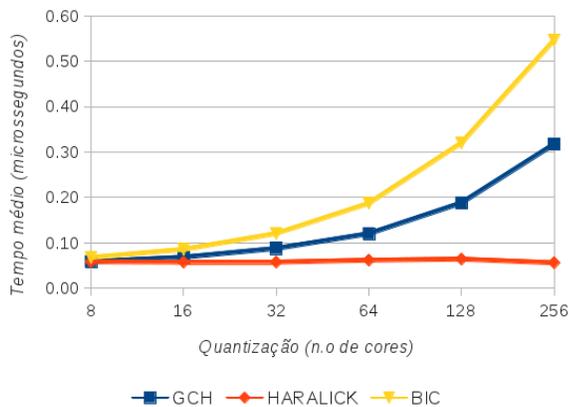


Figura 5. Tempo médio por imagem (em microssegundos) necessário para computar cada um dos descritores utilizando as bases de imagens desse trabalho

Através da análise dos resultados, obtêm-se alguns pontos importantes para este estudo. Entre eles, observa-se que nas bases Corel e Paintings, os descritores BIC e GCH não foram drasticamente afetados com a diminuição da quantidade de cores; a partir de 16 cores os resultados melhoram lentamente. O descritor Haralick produziu melhores resultados com menores quantidades de cores, diferentemente do BIC e do GCH. As evidências apontam que o fato do Haralick ser um descritor de textura e sensível à quantidade de cores está relacionado com a possibilidade de o descritor confundir regiões com muitas cores com informações pertencentes à textura.

Contrariando o padrão de resultados verificado para as bases Corel e Paintings, os experimentos na base Covers produziram resultados confusos e baixa acurácia. Para essa base, será preciso fazer um estudo melhor para verificar a causa da baixa acurácia realizando: a) coleta de mais dados, b) pré-processamento das imagens, c) novo estudo para escolha de descritores.

A análise dos resultados indica que é possível diminuir a quantidade de cores utilizadas na quantização sem que ocorram perdas significativas para a classificação. Esse resultado representa uma importante contribuição pois diversos trabalhos na literatura citados anteriormente apontam de forma arbitrária o uso de valores entre 64 e 256 para o número de cores

Essa diminuição é importante em aplicações como de sistemas embarcados e dispositivos móveis, que possuem a necessidade da utilização de descritores compactos e de classificação em tempo real. A diminuição também se mostra benéfica em aplicações que possuem uma quantidade relativamente grande de imagens, pois com descritores compactos o processamento para extraí-los e analisá-los é reduzido.

Entretanto, a natureza da base de imagem pode influenciar o desempenho da classificação. É necessário estudar as características das imagens da base para avaliar a quantidade de objetos que cada classe deve ter e o nível de quantização a ser utilizado; esses pontos podem determinar o sucesso ou não

de uma aplicação em relação ao desempenho e à acurácia da classificação.

Trabalhos futuros podem estudar o impacto do uso de diferentes métodos de quantização, a relação do melhor nível de quantização para bases de imagens com características gerais diferentes, o desempenho de outros descritores para os diferentes níveis de quantização propostos neste artigo, a relação entre a quantização e a resolução da imagem, bem como aplicações em dispositivos móveis utilizando construídas a partir deste estudo.

#### AGRADECIMENTOS

Esse trabalho foi parcialmente financiado pela FAPESP (proc. n. 2011/16411-4). Agradecimentos ao CNPq pela bolsa de iniciação científica e tecnológica e à USP pelo apoio financeiro.

#### REFERÊNCIAS

- [1] C. Picon, I. Rossi, and M. P. Ponti-Jr, "Análise da classificação de imagens por descritores de cor utilizando várias resoluções," in *SIBGRAP 2011 – Workshop of Undergraduate Works (WUW)*, 2011.
- [2] M. P. Ponti-Jr, "Segmentation of low-cost remote sensing images combining vegetation indices and mean shift," *IEEE Geoscience and Remote Sensing Letters*, 2012, (early access).
- [3] R. Socher and L. Fei-Fei, "Towards total scene understanding: Classification, annotation and segmentation in an automatic framework," in *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 2036–2043.
- [4] O. Penatti, E. Valle, and R. Torres, "Comparative study of global color and texture descriptors for web image retrieval," *Journal of Visual Communication and Image Representation*, vol. 23, no. 2, pp. 359–380, 2012.
- [5] K. Choi, K.-A. Toh, and H. Byun, "Realtime training on mobile devices for face recognition applications," *Pattern Recognition*, vol. 44, no. 2, pp. 386–400, 2011.
- [6] H. Yang, A. Kot, and X. Jiang, "Dynamic window construction for the binarization of barcode images captured by mobile phones," in *17th Int. Conf. Image Processing (ICIP)*. IEEE, 2010, pp. 3001–3004.
- [7] L. Gorlenko and R. Merrick, "No wires attached: usability challenges in the connected mobile world," *IBM Systems Journal*, vol. 42, no. 4, pp. 639–651, 2003.
- [8] M. J. Swain and D. H. Ballard, "Color indexing," *Int. J. Computer Vision*, vol. 7, no. 11, pp. 11–32, 1991.
- [9] R. O. Stehling, M. A. Nascimento, and A. X. Falcao, "A compact and efficient image retrieval approach based on border/interior pixel classification," in *CIKM'02 - 11th ACM Int Conf Information Knowledge Management*, 2002, pp. 102–109.
- [10] R. Haralick, K. Shanmugan, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.
- [11] J. Papa, A. X. Falcao, and C. T. N. Suzuki, "Supervised pattern classification based on optimum-path forest," *Int. J. Imaging Systems and Technology*, vol. 19, no. 2, pp. 120–131, 2009.
- [12] M. P. Ponti-Jr and J. P. Papa, "Improving accuracy and speed of optimum-path forest classifier using combination of disjoint training subsets," in *10th Int. Work. on Multiple Classifier Systems (MCS 2011) LNCS 6713*. Naples, Italy: Springer, 2011, pp. 237–248.
- [13] M. P. Ponti-Jr, J. P. Papa, and A. L. M. Levada, "A Markov Random Field model for combining Optimum-Path Forest classifiers using decision graphs and game strategy approach," in *16th Iberoamerican Congress on Pattern Recognition (CIARP 2011)*, vol. LNCS 7042. Pucon, Chile: Springer, 2011, pp. 581–590.
- [14] J. Li and J. Z. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1075–1088, 2003.
- [15] Intel, *OpenCV: open source computer vision library*, 2nd ed., Intel Corporation, <http://opencv.willowgarage.com/wiki/>, Abril 2010, BSD license.