

# NOVAS CARACTERÍSTICAS PARA CLASSIFICAÇÃO DE FORMAS INVARIANTE SOB TRANSFORMAÇÃO AFIM

CARLOS R. P. DIONISIO\*, HAE YONG KIM\*

\**Departamento de Engenharia de Sistemas Eletrônicos, Escola Politécnica,  
Universidade de São Paulo, São Paulo, SP, Brasil*

Emails: carlos@lps.usp.br, hae@lps.usp.br

**Abstract**— An object seen from different viewpoints results in differently deformed images. Affine-invariant shape classification must classify correctly the object, regardless its viewpoint. In this paper, we propose new local and global features invariant under affine transformation. These features can be used for the supervised or unsupervised shape classification, and for the shape-based image indexing and retrieval. One of the proposed features is related to the convex deficiency and the others are extracted from the area matrix. Area matrix was used by Shen (Shen et al., 1999) for the similarity matching in the image retrieval. However, differently from the Shen's work, we parameterize the shape contour using the affine-length parameter. This makes our features robust to shape parameterization, while Shen's work does not have this property. Experimental results indicate that our method can classify correctly even highly deformed and noisy shapes using a small training set.

**Keywords**— shape classification; image retrieval; affine-length; affine transformation.

**Resumo**— Um objeto visto de diferentes pontos de vista resulta em imagens diferentemente deformadas. A classificação de formas invariante afim deve classificar corretamente um objeto, independentemente do ponto de vista. Neste trabalho, propomos novas características locais e globais invariante sob transformação afim. Estas características podem ser usadas para a classificação supervisionada ou não-supervisionada de formas, e para a recuperação e indexação de imagens baseado em forma. Uma das características propostas está relacionada com a deficiência convexa e as outras são extraídas da matriz de áreas. A matriz de áreas foi usada por Shen (Shen et al., 1999) para o casamento por similaridade na recuperação de imagens. Porém, diferentemente do trabalho de Shen, parametrizamos os contornos das formas usando o parâmetro comprimento de arco afim. Isto torna as nossas características robusto a parametrização da forma, enquanto que o trabalho de Shen não possui esta propriedade. Os resultados experimentais indicam que o nosso método pode classificar corretamente até as formas altamente deformadas e ruidosas usando um pequeno conjunto de treinamento.

**Keywords**— classificação de formas; recuperação de imagens; comprimento de arco afim; transformação afim.

## 1 Introdução

Um problema importante na classificação de objetos é o fato de que um objeto pode ser visto desde diferentes pontos de vista, resultando em imagens diferentes. Conseqüentemente, a invariância a ponto de vista é uma propriedade desejável em muitos sistemas de reconhecimento de formas. Para objetos quase planos, estas deformações podem ser modeladas aproximadamente por transformações afins, se os pontos de vista estiverem suficientemente distantes do objeto. Na literatura, existem muitos trabalhos sobre o reconhecimento de formas invariante sob transformação afim. Alguns exemplos são os descritores de Fourier invariantes afins (Arbter et al., 1990), os momentos invariantes afins (Flusser and Suk, 1993; Huang and Cohen, 1994; Zhao and Chen, 1997) e o espaço de escala curvatura afim (Mokhtarian and Abbasi, 2001). Nestes trabalhos, a idéia fundamental é usar uma parametrização afim, usualmente o comprimento de arco afim (Arbter et al., 1990; Flusser and Suk, 1993; Zhao and Chen, 1997; Sapiro and Tannenbaum, 1993).

Usualmente, uma boa taxa de classificação pode ser obtida usando simultaneamente as características locais e globais. As características locais são sensíveis a ruídos e, na prática, a quantidade de informações locais disponíveis são

usualmente insuficientes para uma classificação robusta. Alguns exemplos são os invariantes geométricos (Rivlin and Weiss, 1995), as esquinas (corners) (Rattarangsi and Chin, 1992) e as curvaturas (Mokhtarian, 1995). Por outro lado, as características globais são relativamente insensíveis a ruídos e duas formas similares normalmente possuem as características globais semelhantes. Dois exemplos simples são a área e a deficiência convexa.

A deficiência convexa é uma característica global que contém informação sobre as concavidades da forma. Esta característica é invariante por translação, rotação e facilmente pode se tornar invariante à escala. Neste trabalho, mostramos que a deficiência convexa pode também se tornar invariante à transformação afim. As outras características propostas são extraídas da matriz de áreas. A matriz de áreas foi usada por Shen (Shen et al., 1999) para o casamento de formas por similaridade em recuperação de imagens invariante afim. Esta matriz possui duas propriedades interessantes. A primeira é que ela pode se tornar invariante à transformação afim. A segunda é que as suas colunas contém informações multi-escala: as colunas centrais contém informações globais ou em escala grossa e as colunas periféricas contém informações locais ou em escala fina. Diferentemente do trabalho de Shen, parametrizamos as

formas usando o parâmetro comprimento de arco afim. Este procedimento torna a matriz de áreas invariante sob parametrização afim. A complexidade computacional do trabalho de Shen depende da dimensão da matriz de áreas. Pelo contrário, extraímos e processamos somente as informações relevantes, diminuindo a complexidade computacional. Na fase de reconhecimento, usamos o algoritmo dos  $k$ -vizinhos mais próximos.

O resto do trabalho está organizado como segue. A Seção 2 trata da área, da transformação afim e da parametrização afim. Na Seção 3, descrevemos a matriz de áreas, a sua normalização e a extração de informações relevantes. A Seção 4 trata das características propostas, do algoritmo de treinamento e do algoritmo de classificação. Os resultados experimentais estão expostos na Seção 5. Finalmente, na Seção 6 apresentamos as conclusões deste trabalho.

## 2 Conceitos Fundamentais

### 2.1 Curva Fechada

Uma forma plana pode ser definida pelo seu contorno. Neste trabalho, trataremos somente das curvas fechadas no plano, supondo que estas curvas são os contornos das formas planas.

Consideremos a equação paramétrica discreta de uma curva fechada:

$$\mathbf{u}(t) = (x(t), y(t)) \quad (1)$$

onde  $t \in \{0, \dots, N-1\}$  e  $\mathbf{u}(N) = \mathbf{u}(0)$ . Em geral, as curvas de uma aplicação podem ter sido parametrizadas com um número diferente de vértices  $N$ .

A primeira e segunda derivadas de  $\mathbf{u}(t)$  são denotadas, respectivamente, como  $\dot{\mathbf{u}}(t) = (\dot{x}(t), \dot{y}(t))$  e  $\ddot{\mathbf{u}}(t) = (\ddot{x}(t), \ddot{y}(t))$ .

### 2.2 Transformação Afim

Uma transformação afim geral  $T$  em  $\mathbb{R}^2$  é definida como:

$$T(X) = AX + b \quad (2)$$

onde  $X \in \mathbb{R}^2$ ,  $A$  é a matriz afim (uma matriz real  $2 \times 2$  com determinante positivo) e  $b \in \mathbb{R}^2$  é o vetor translação. Uma transformação afim geral contém as transformações de translação, escala, rotação e deformação (shearing). A seguinte matriz representa uma transformação de deformação:

$$A_\kappa = \begin{bmatrix} 1 & \kappa \\ 0 & 1 \end{bmatrix} \quad (3)$$

onde o parâmetro  $\kappa$ , chamado de taxa de deformação (shear ratio), controla a medida da deformação. As transformações de rotação, escala uniforme e translação não alteram a forma. Por outro lado, as transformações de escala não-uniforme e

deformação modificam a forma. Neste trabalho, testamos a robustez das nossas características sob transformações de rotação e deformação.

### 2.3 Área e Transformação Afim

Seja  $\Delta$  o triângulo formado pelos pontos  $(x_1, y_1)$ ,  $(x_2, y_2)$  e  $(x_3, y_3)$ . Então, a área do triângulo  $\Delta$  é dada pela seguinte fórmula:

$$\text{Area}(\Delta) = \frac{1}{2} \det \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{bmatrix} \quad (4)$$

Se  $T$  é uma transformação afim e  $A$  a sua matriz, então:

$$\text{Area}(T(\Delta)) = |A| \text{Area}(\Delta) \quad (5)$$

onde  $|A|$  é o determinante de  $A$ . Considerando a Equação 5 e o fato de que qualquer polígono simples 2-D  $\mathbf{u}$  pode ser descomposto como um conjunto de triângulos, podemos afirmar o seguinte:

$$\text{Area}(T(\mathbf{u})) = |A| \text{Area}(\mathbf{u}) \quad (6)$$

Seja  $\text{FC}(\mathbf{u})$  o fecho convexo da curva  $\mathbf{u}$ . Claramente, um ponto pertence ao  $\text{FC}(\mathbf{u})$  se, e somente se, a sua transformada pertencer ao  $\text{FC}(T(\mathbf{u}))$ . Conseqüentemente, podemos também afirmar o seguinte:

$$\text{Area}(\text{FC}(T(\mathbf{u}))) = |A| \text{Area}(\text{FC}(\mathbf{u})) \quad (7)$$

Baseado nas Equações 6 e 7, definimos a nossa primeira característica invariante afim, dividindo a deficiência convexa de  $\mathbf{u}$  pela área do fecho convexo de  $\mathbf{u}$ :

$$F_1 = \frac{\text{Area}(\text{FC}(\mathbf{u})) - \text{Area}(\mathbf{u})}{\text{Area}(\text{FC}(\mathbf{u}))} \quad (8)$$

### 2.4 Parametrização Afim

O comprimento de arco de uma curva  $\mathbf{u}(t) = (x(t), y(t))$  pode ser calculado usando a seguinte fórmula:

$$s(p) = \frac{\int_0^p (\dot{x}^2(t) + \dot{y}^2(t))^{\frac{1}{2}} dt}{\int_0^N (\dot{x}^2(t) + \dot{y}^2(t))^{\frac{1}{2}} dt} \quad (9)$$

Claramente, o comprimento de arco  $s$  não é linear sob transformação afim, ao contrário do parâmetro comprimento de arco afim definido abaixo (Arbter et al., 1990; Flusser and Suk, 1993; Zhao and Chen, 1997; Sapiro and Tannenbaum, 1993):

$$\tau(p) = \frac{\int_0^p |\dot{x}(t)\ddot{y}(t) - \ddot{x}(t)\dot{y}(t)|^{\frac{1}{3}} dt}{\int_0^N |\dot{x}(t)\ddot{y}(t) - \ddot{x}(t)\dot{y}(t)|^{\frac{1}{3}} dt} \quad (10)$$

Para garantir que  $\tau$  seja uma função estritamente crescente, removemos todos os vértices com a segunda derivada nula. Isto equivale a remover os vértices pertencentes a linhas retas.

Neste trabalho, normalizamos o número de vértices de todas as formas, pois as características propostas dependem da escolha deste número. Também re-parametrizamos todas as formas usando o parâmetro comprimento de arco afim. Desta forma, as nossas características serão robustas a parametrização da forma pelo comprimento de arco afim.

### 3 Extração de Informações Locais e Globais de uma Matriz de Áreas

#### 3.1 Matriz de Áreas

Seja  $\mathbf{u}(t)$ ,  $t \in \{0, \dots, N-1\}$ , uma curva fechada e seja  $m = N/2$ . Seja  $\mathbf{A}_{ti}$  a área do triângulo formado pelos vértices  $\mathbf{u}(t-i)$ ,  $\mathbf{u}(t)$  e  $\mathbf{u}(t+i)$ . Então, a matriz de áreas de  $\mathbf{u}(t)$  é definida como:

$$\mathbf{A} = [\mathbf{A}_{ti}]_{i=1, \dots, m}^{t=0, \dots, N-1} \quad (11)$$

O  $i$ -ésimo vetor coluna de  $\mathbf{A}$  contém informações locais (escala fina) quando o índice  $i$  for pequeno ou próximo a  $m$  e contém informações globais (escala grossa) quando  $i$  estiver situado na região central de  $\mathbf{A}$ . De fato, cada vetor coluna da matriz de áreas analisa a forma numa determinada escala, e a matriz de áreas completa contém as informações multi-escala da forma.

#### 3.2 Matriz de Áreas Normalizada

Normalizamos a matriz de áreas  $\mathbf{A}$  em dois passos. Primeiro, tornamos a matriz  $\mathbf{A}$  invariante à transformação afim dividindo cada elemento de  $\mathbf{A}$  pela soma de todos os elementos de  $\mathbf{A}$ . Segundo, fazemos todas as colunas terem magnitudes similares: Dividimos cada coluna  $i$  de  $\mathbf{A}$  pelo peso  $w(i)$ . O vetor de pesos  $w$  é calculado a partir de uma matriz de áreas invariante afim  $\mathbf{R}$  de uma forma representativa usando a seguinte equação:

$$w(i) = \max_{t=0, \dots, N-1} \{|\mathbf{R}_{t,i}|\} \quad i = 1, \dots, m \quad (12)$$

#### 3.3 Extração de Informações Locais e Globais

Uma matriz de áreas contém informações em quantidade excessiva. Para ser útil, ela precisa ser reduzida e organizada de alguma forma. Extraímos as informações relevantes a partir de uma matriz de áreas normalizada  $\mathbf{A}$  e as armazenamos em três vetores  $\psi_e$ ,  $\psi_c$  e  $\psi_d$  que resumem respectivamente as informações das colunas situadas à esquerda, no centro e à direita da matriz  $\mathbf{A}$ . O vetor  $\psi_e$  é calculado da seguinte forma:

$$\psi_e(t) = \frac{\sum_i g_e(i) \mathbf{A}_{t,i}}{\sum_i g_e(i)} \quad t = 0, \dots, N-1 \quad (13)$$

onde  $g_e$  é uma curva gaussiana. O centro desta gaussiana está situado no lado esquerdo de  $\mathbf{A}$ . Ele não pode estar muito próximo à primeira coluna,

para reduzir os ruídos presentes em escala finas. O desvio-padrão  $\sigma$  é escolhido para extrair todas as informações presentes no lado esquerdo de  $\mathbf{A}$ , sem que haja mistura com as informações da região central de  $\mathbf{A}$ . Os vetores  $\psi_c$  e  $\psi_d$  são calculados de forma similar.

Neste trabalho, as três gaussianas foram definidas de acordo a Tabela 1:

Tabela 1: Configuração das três gaussianas ( $m$  é o número de colunas de  $\mathbf{A}$ ).

Gaussiana	$\sigma$	centro
$g_e$	5	0.20m
$g_c$	5	0.50m
$g_d$	5	0.85m

## 4 Classificação Supervisionada

Um dos principais problemas da Visão Computacional é a classificação supervisionada de objetos. Neste problema, é necessário dispor de um *conjunto de exemplos*, previamente conhecido e particionado em classes. Este conjunto é também conhecido como o *conjunto de treinamento*. A classificação supervisionada compreende duas fases: *aprendizado* e *reconhecimento*. A fase de aprendizado corresponde ao estágio onde os critérios e os métodos de classificação são construídos, usando os dados do conjunto de treinamento. A atribuição de uma classe a um objeto de consulta corresponde à fase de reconhecimento.

#### 4.1 Características Propostas

Os vetores  $\psi_e$ ,  $\psi_c$  e  $\psi_d$  são definidos como na Subseção 3.3. Usamos cinco características para classificar as formas:

- A característica  $F_1$  foi definida na Equação 8.
- A característica  $F_2$  é a soma dos valores negativos de  $\psi_e$ , isto é,  $F_2 = \sum_{\psi_e(t) < 0} \psi_e(t)$ .
- A característica  $F_3$  é a soma dos valores positivos de  $\psi_e$ , isto é,  $F_3 = \sum_{\psi_e(t) > 0} \psi_e(t)$ .
- A característica  $F_4$  é a soma de  $\psi_c$ , isto é,  $F_4 = \sum \psi_c(t)$ .
- A característica  $F_5$  é a soma de  $\psi_d$ , isto é,  $F_5 = \sum \psi_d(t)$ .

Todas estas características são invariantes sob transformação afim. Elas também são invariantes à escolha do vértice inicial. Para poder usar o algoritmo de classificação dos  $k$ -vizinhos mais próximos, normalizamos essas características para terem a média zero e o desvio-padrão um.

## 4.2 Algoritmo de Treinamento

O algoritmo de treinamento consiste em calcular as cinco características e deve ser aplicado a todos os exemplos de treinamento. Cada exemplo de treinamento deve ser apresentado a este algoritmo junto com a sua correta classificação.

- Normalize o número de vértices da forma.
- Re-parametrize a forma usando o parâmetro comprimento de arco afim (Equação 10).
- Calcule e normalize a matriz de áreas.
- Determine os vetores  $\psi_e$ ,  $\psi_c$  e  $\psi_d$ .
- Calcule as cinco características da forma.

## 4.3 Algoritmo de Classificação

O algoritmo de classificação recebe como entrada uma forma de consulta com classificação desconhecida e o número  $k$  a ser usado no algoritmo dos  $k$ -vizinhos mais próximos. A saída deste algoritmo é a classificação da forma.

- Dada uma forma de consulta, calcule as cinco características como descrito no algoritmo de treinamento. Denotemos o vetor de características obtido como  $x$ .
- Procure, no conjunto de treinamento,  $k$  formas que estão situadas mais próximas ao vetor  $x$ . A classificação será a moda das classificações das  $k$  formas de treinamento (isto é, a classificação mais freqüente).

## 5 Resultados Experimentais

Aplicamos o nosso método para a classificação supervisionada dos gestos estáticos de mão. Usamos parte dos contornos dos gestos estáticos de mão considerados por Milios e Petrakis (Milios and Petrakis, 2000). Todos os contornos foram normalizados para terem 140 vértices. Rotulamos manualmente 150 contornos fechados como exemplos de treinamento supervisionado. Eles foram divididos em seis classes, cada uma com 25 exemplos, conforme ilustrado na Figura 1.

### 5.1 Formas Deformadas

Testamos a invariância afim das nossas características, especificamente para as transformações de rotação e deformação. Tomamos um contorno representativo de cada classe (que não pertence ao conjunto de treinamento) e geramos 9 versões rotacionadas, rodando o contorno original por  $20^\circ$ ,  $40^\circ$ , ...,  $180^\circ$ . Depois, deformamos as  $6 \times 9$  versões rotacionadas com as taxas de deformação  $\kappa = 1, 2$  e  $3$ , obtendo  $3 \times 6 \times 9$  formas deformadas. As Figuras 2, 3 e 4 ilustram a forma original e as versões

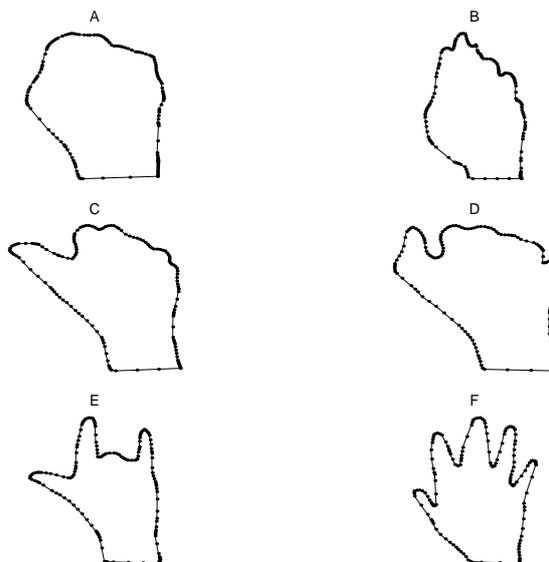


Figura 1: Gestos de mão das classes A, B, C, D, E e F.

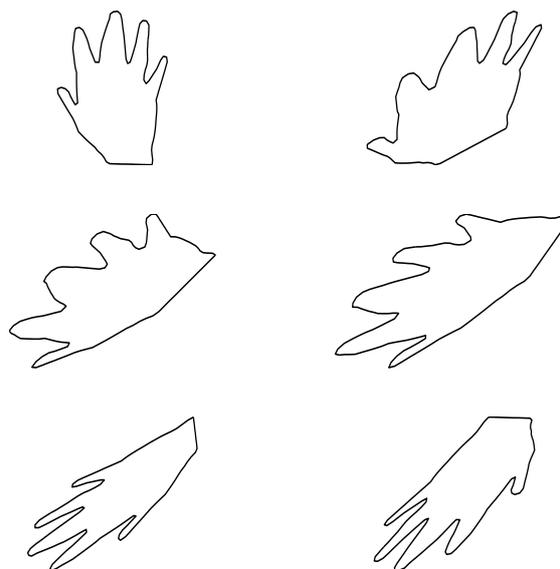


Figura 2: O gesto na parte superior esquerda pertence à classe F. Os outros são as formas deformadas com  $\kappa = 1$ .

deformadas com as taxas de deformação  $\kappa = 1, 2$  e  $3$ , respectivamente.

Para a fase de reconhecimento, usamos o algoritmo dos  $k$ -vizinhos mais próximos. Testamos este algoritmo com  $k$  variando de 1 até 25. Todas as formas deformadas foram corretamente classificadas, para qualquer  $k \in \{1, \dots, 25\}$ . Isto mostra a robustez das características propostas sob transformação afim.

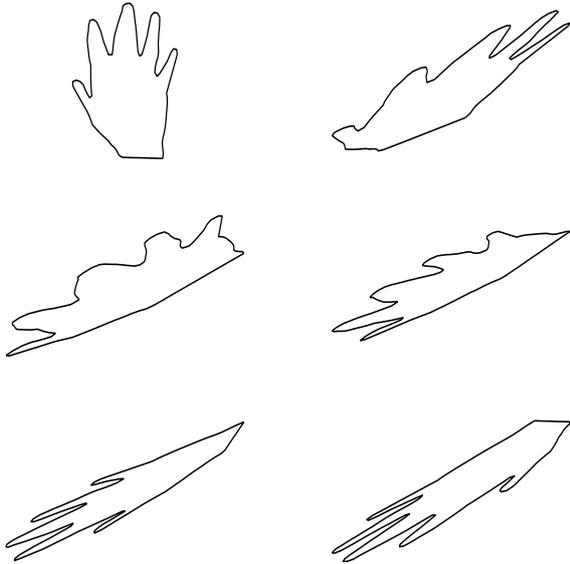


Figura 3: O gesto na parte superior esquerda pertence à classe F. Os outros são as formas deformadas com  $\kappa = 2$ .

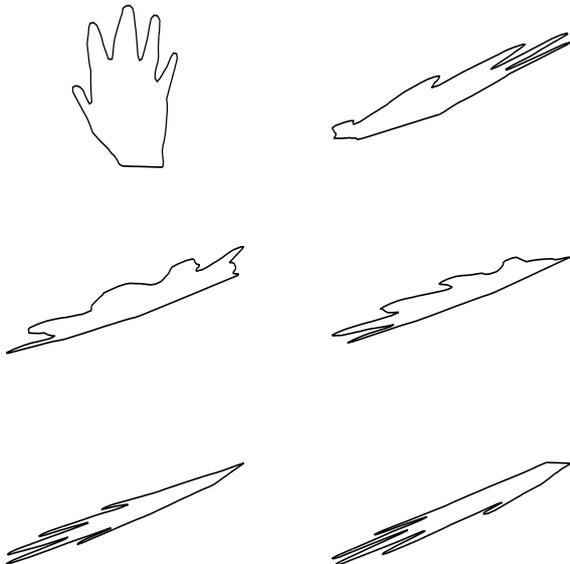


Figura 4: O gesto na parte superior esquerda pertence à classe F. Os outros são as formas deformadas com  $\kappa = 3$ .

### 5.2 Formas Não-Deformadas

Também rotulamos manualmente 150 contornos de consulta não-deformados, com 25 contornos em cada classe. O conjunto dos contornos de consulta é disjunto do conjunto dos contornos de treinamento. A Tabela 2 resume a classificação dos gestos de mão de consulta para as formas não-deformadas, usando o algoritmo dos 5-vizinhos mais próximos. Nesta tabela, o elemento (linha  $i$ , coluna  $j$ ) corresponde ao número de gestos de mão manualmente rotulados como  $i$  e classificados pelo algoritmo como  $j$ . A taxa de erro é muito

Tabela 2: Classificação dos contornos de consulta.

	A	B	C	D	E	F
A	25	0	0	0	0	0
B	1	24	0	0	0	0
C	0	0	25	0	0	0
D	0	1	0	24	0	0
E	0	0	0	0	25	0
F	0	0	0	1	0	25

baixa, o que mostra a aptidão das características propostas para a classificação de formas. As Figuras 5, 6 e 7 ilustram alguns contornos de consultas não-deformados e as suas classificações.

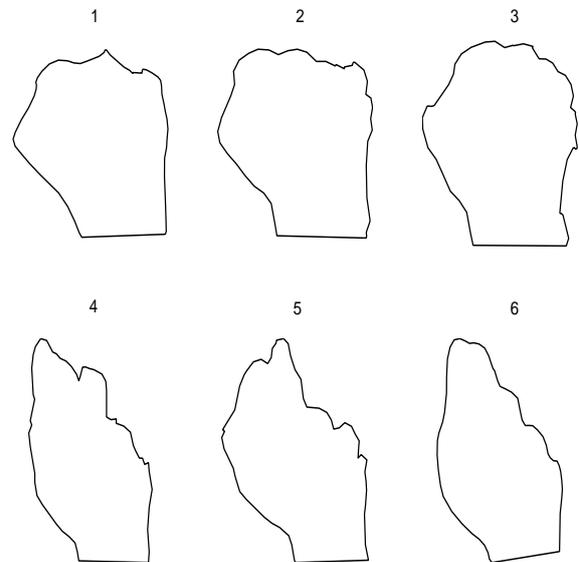


Figura 5: Os gestos 1, 2 e 3 pertencem à classe A e eles foram corretamente classificados pelo nosso algoritmo. Os gestos 4, 5 e 6 pertencem à classe B. Os gestos 4 e 5 foram corretamente classificados, mas o gesto 6 foi classificado como A.

## 6 Conclusões

Neste trabalho, propusemos um conjunto de novas características invariante sob transformação afim. Utilizamos estas características para a classificação supervisionada de formas, mas elas podem também ser usadas na classificação não-supervisionada de formas e na indexação e recuperação de imagens baseado em forma. Parametrizamos as formas usando o parâmetro comprimento de arco afim, para tornar as nossas características robusto a parametrização da forma. Os resultados experimentais mostraram que as nossas características são capazes de discriminar corretamente formas severamente deformadas sob transformação afim.

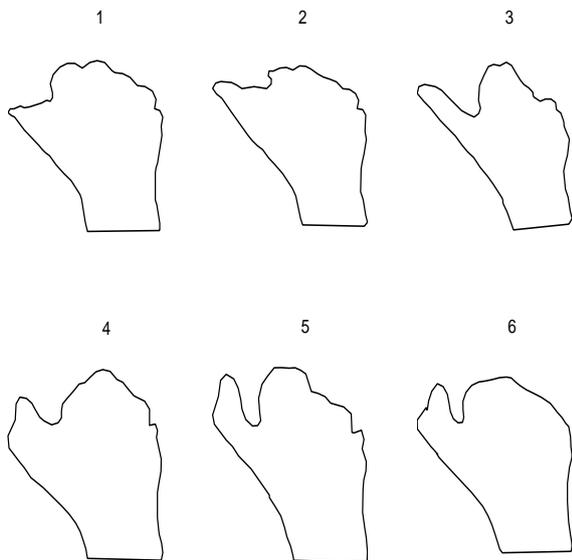


Figura 6: Os gestos 1, 2 e 3 pertencem à classe C e eles foram corretamente classificados pelo nosso algoritmo. Os gestos 4, 5 e 6 pertencem à classe D. Os gestos 4 e 5 foram corretamente classificados, mas o gesto 6 foi classificado como B.

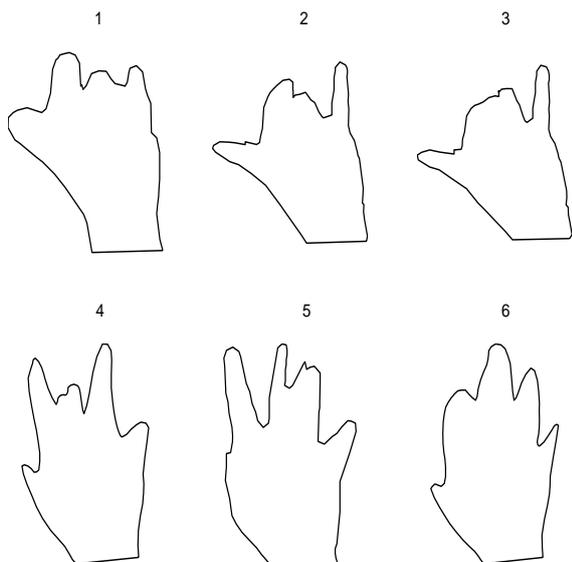


Figura 7: Os gestos 1, 2 e 3 pertencem à classe E e eles foram corretamente classificados pelo nosso algoritmo. Os gestos 4, 5 e 6 pertencem à classe F. Os gestos 4 e 5 foram corretamente classificados, mas o gesto 6 foi classificado como D.

## Agradecimentos

Carlos R. P. Dionisio gostaria de agradecer à Capes pela bolsa de doutorado que financiou este trabalho. Hae Yong Kim gostaria de agradecer ao CNPq pelo apoio financeiro deste trabalho sob processo 300689/98-5.

## Referências

- Arbter, K., Snyder, W. and H. Burkhardt, G. H. (1990). Applications of affine-invariant fourier descriptors to recognition of 3-d objects, *IEEE Trans. Pattern Analysis and Machine Intelligence* **12**(7): 640–646.
- Flusser, J. and Suk, T. (1993). Pattern recognition by affine moment invariants, *Pattern Recognition* **26**(1): 167–174.
- Huang, Z. and Cohen, F. (1994). Affine-invariant b-spline moment for curve matching, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 490–495.
- Milios, E. and Petrakis, E. (2000). Shape retrieval based on dynamic programming, *IEEE Transactions on Image Processing* **9**(1): 141–146.
- Mokhtarian, F. (1995). Silhouette-based isolated object recognition through curvature-scale space, *IEEE Trans. PAMI* **17**(5): 539–544.
- Mokhtarian, F. and Abbasi, S. (2001). Affine curvature scale space with affine length parametrisation, *Pattern Analysis & Applications* **4**: 1–8.
- Rattarangsi, A. and Chin, R. (1992). Scale-based detection of corners of planar curves, *IEEE Trans. PAMI* **14**(4): 430–449.
- Rivlin, E. and Weiss, I. (1995). Local invariants for recognition, *IEEE Trans. PAMI* **17**(3): 226–238.
- Sapiro, S. and Tannenbaum, A. (1993). Affine invariant scale-space, *International Journal of Computer Vision* **11**(1): 25–44.
- Shen, D., Wong, W. and Ip, H. (1999). Affine-invariant image retrieval by correspondence matching of shapes, *Image and Vision Computing* **17**: 489–499.
- Zhao, A. and Chen, J. (1997). Affine curve moment invariants for shape recognition, *Pattern Recognition* **30**(6): 895–901.