

## 5 Reconhecimento

No decorrer deste capítulo é proposto um modelo para o reconhecimento de gestos da mão baseado na silhueta. Além dos gestos, são detectadas a posição e a orientação dos dedos. Num primeiro passo são localizadas algumas feições na silhueta, depois essas feições são avaliadas para reconhecer os dedos e finalmente alguns gestos pré-estabelecidos são reconhecidos. Na próxima seção é apresentada uma motivação geral do reconhecimento, na qual são definidos alguns parâmetros de interesse. Em seguida é apresentado um breve resumo do modelo e finalmente, de forma mais detalhada, cada um de seus passos, os resultados obtidos e algumas considerações finais.

### 5.1. Motivação

Os dois capítulos anteriores foram focados na localização de regiões de interesse nas imagens analisadas. O próximo passo é observar mais de perto essas regiões e encontrar algumas características que permitam extrair informações relevantes acerca da posição dos dedos e do gesto que está sendo mostrado.

Existem dois tipos de gestos que poderiam ser reconhecidos na mão. Os gestos estáticos são aqueles que mostram apenas uma determinada postura da mão; o significado está na forma ou postura que apresentam. Os gestos dinâmicos são aqueles em que seu significado depende também do movimento que a mão descreve. Nos gestos estáticos é importante apenas saber qual é o gesto da mão a cada quadro, independentemente do tempo ou do número de quadros. Nos gestos dinâmicos é importante analisar o comportamento da mão ao longo do tempo. Certamente os gestos dinâmicos podem ser reconhecidos ou inferidos a partir de um conjunto de gestos estáticos. Por exemplo, a detecção de movimento poderia ser reconhecida a partir de um conjunto de gestos estáticos ou do comportamento dos gestos estáticos em um certo período de tempo. No presente trabalho o objetivo desta etapa é construir um módulo para o reconhecimento de alguns gestos básicos. As aplicações utilizam tais gestos

básicos de forma particular e de acordo com suas necessidades. Portanto, nesta etapa são compreendidos apenas os gestos estáticos, os quais serão denominados no texto simplesmente *gestos*.

O reconhecimento dos gestos é feito com base no número de dedos que em um determinado momento a silhueta da mão apresenta. Portanto, para reconhecer os gestos é importante determinar com clareza o número de dedos presentes e também suas características, tais como sua posição, orientação e os pontos que os representam na silhueta. Levando em conta os dedos, existe um grande número de gestos que podem ser construídos, mas serão apenas 4 os gestos básicos reconhecidos neste trabalho. Esses gestos tentam ser os mais básicos possíveis, de forma que eles possam servir para construir ou reconhecer tipos de gestos mais complicados nas aplicações.

Na literatura apresentada no segundo capítulo vimos que não existe uma única abordagem para reconhecer gestos da mão em imagens. Algumas técnicas utilizam forma, outras forma e cor, e outras utilizam ainda transformadas para o reconhecimento de gestos da mão. Apesar dessa diversidade, existem alguns parâmetros de interesse que são considerados pela maioria dessas abordagens. Entre eles podemos citar:

- Número de dedos presentes na mão a cada quadro.
- Posição e orientação dos dedos presentes com relação à imagem.
- Posição e orientação dos dedos com relação à mão.
- Posição e orientação da mão.
- Posição 3D dos pontos que representam os dedos.

Os parâmetros considerados mais importantes neste trabalho são os três primeiros. Descartamos as abordagens tridimensionais por serem computacionalmente caras. Além disso, o reconhecimento 3D é condicionado e direcionado a certo tipo específico de aplicações.

O processo de reconhecimento de gestos apresentado aqui possui três passos básicos. No primeiro passo a silhueta é analisada para procurar pontos que apresentem características dos dedos. Essa procura se baseia em medidas de curvatura que indicam alguns pontos candidatos. Alguns cuidados são levados em conta de maneira que o processo de detecção de feições seja dirigido a procurar, em sua maioria, os pontos que representam os dedos.

No segundo passo esses pontos são analisados de maneira a selecionar aqueles pontos que melhor se assemelhem às características que procuramos nos dedos. Além disso, as posições e as orientações dos dedos são calculadas.

No último passo, os gestos são inferidos levando em conta o número e a característica dos dedos. Esses gestos são selecionados de um conjunto de gestos pré-estabelecidos para o reconhecimento. Nas próximas seções é apresentado em mais detalhe cada um dos passos do processo de reconhecimento de gestos.

## 5.2. Detecção de Feições

Na silhueta da mão existem duas feições a serem consideradas para o reconhecimento dos gestos (Figura 26): as pontas dos dedos, chamadas também de Picos (“P”), e as junções entre os dedos, denominados Vales (“V”). Identificar as regiões da silhueta que contêm esses pontos é importante para que o reconhecimento possa ser realizado satisfatoriamente.

O objetivo desta etapa é identificar na silhueta da mão os conjuntos de pontos que mais se assemelhem com os Picos e Vales. Nesse processo não são encontrados diretamente os Picos nem os Vales, mas sim conjuntos de pontos com características similares nos quais também estão incluídos os Picos e Vales.

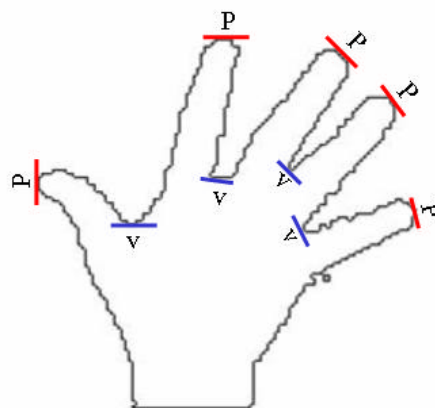


Figura 26 - Feições consideradas para o reconhecimento dos gestos. As pontas dos dedos são chamadas de Picos (“P”) e as junções entre eles de Vales (“V”).

O problema de se localizar os Picos e Vales na silhueta de uma mão pode ser reduzido ao problema de se encontrar cantos (*corners*) em curvas planares. Estes cantos, dentro da curva que descreve a mão, conceitualmente podem ser considerados como sendo máximos locais (para os Picos) e mínimos locais

(para os Vales). Em nosso caso as curvas possuem muito ruído, como ilustra a Figura 26. O ruído faz com que as medidas de curvatura, baseadas em derivadas locais, não sejam adequadas. Muitos pontos ao longo da curva têm alto valor de curvatura e seriam erroneamente considerados como pontos de interesse.

O método utilizado para encontrar os Picos e Vales na silhueta deve ser insensível aos ruídos sem perder os verdadeiros pontos de interesse. A literatura apresenta diversos algoritmos para calcular curvaturas na presença de ruídos. Nos algoritmos descritos no trabalho de Liu & Srinath [25], por exemplo, a medida de curvatura é calculada através de um ângulo. Quando um ponto é processado, o valor de curvatura é calculado levando em conta um conjunto de pontos subseqüentes e anteriores a esse ponto na silhueta. Assim, dependendo do número de pontos a serem considerados, essa medida de curvatura pode ser pouco afetada pelos ruídos de alta freqüência presentes ao longo da silhueta.

Apesar desses algoritmos poderem ser utilizados para nossos propósitos, neste trabalho é utilizado um cálculo de curvatura um tanto diferente que considera um número menor de pontos vizinhos na silhueta e será aqui chamado de *k-curvatura*. Quando um ponto  $P_i$  é processado, seu valor de *k-curvatura* é o ângulo  $\alpha_i$  formado pelos vetores  $(P_i^+ - P_i)$  e  $(P_i^- - P_i)$ , ilustrados na Figura 27.(a).

Diferentemente dos algoritmos descritos por Liu & Srinath [17], nos quais vários pontos são considerados para o cálculo da curvatura, em nosso cálculo são levados em conta dois vetores (Figura 27.(a)) formados por um único ponto subseqüente  $P_{(i+k)}$  e outro anterior  $P_{(i-k)}$  ao ponto  $P_i$  na silhueta, sendo  $k$  um inteiro positivo. Observando a Figura 27.(a) pode-se notar que utilizando esses dois vetores a tendência é diminuir o efeito do ruído ao longo da silhueta e ter ângulos menores onde a *k-curvatura* é maior. Em função do valor de  $k$ , do qual depende o tamanho dos vetores, o problema dos ruídos ao longo da silhueta pode afetar pouco o cálculo da *k-curvatura*.

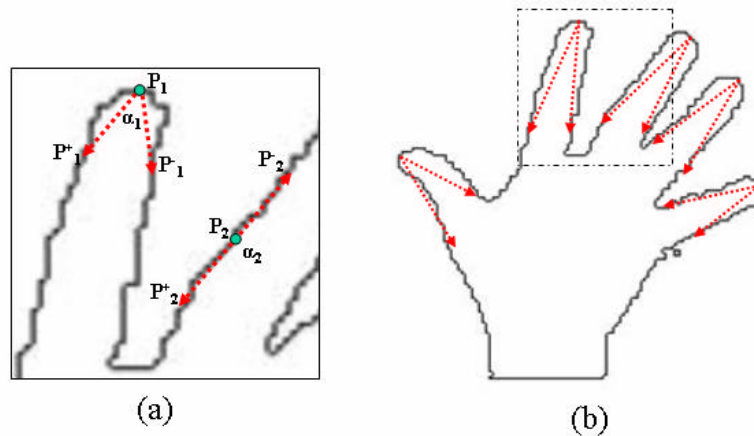


Figura 27 - Medida de k-curvatura em cada ponto da silhueta. (a) Para cada ponto  $P_i$  é escolhido um ponto subsequente  $P_{(i+k)}$  e outro anterior  $P_{(i-k)}$  que formam dois vetores. A k-curvatura é o ângulo  $\alpha_i$  formado por esses dois vetores e  $k$  é um inteiro positivo. (b) O parâmetro  $k$  deve ser ajustado de maneira que as regiões de interesse sejam preferentemente detectadas durante o cálculo da k-curvatura.

Para detectar principalmente as regiões com Picos e Vales é necessário levar em conta algumas observações referentes à forma da mão e explorar as características dos dedos na silhueta. Observando a Figura 27.(b) pode-se ver que os dedos são representados por um número considerável de pontos na silhueta, e esses pontos estão distribuídos de modo a formar regiões cilíndricas ou semi-cápsulas ao longo da silhueta. Assim, acrescentando o valor de  $k$  até um certo valor, é possível formar vetores suficientemente longos de modo que a característica dos dedos estendidos seja aproveitada na detecção dos Picos e Vales. Esses vetores são mostrados Figura 27.(b). É importante atribuir um valor a  $k$  que considere os dedos estendidos, mas que também possa ser flexível quando os dedos não estejam totalmente estendidos, por isso a escolha do valor de  $k$  é importante para o aproveitamento dessa característica na detecção de feições.

No algoritmo de detecção de feições, além da escolha do valor de  $k$ , é necessário considerar um limiar  $T_a$  para o valor da k-curvatura. Quando um ponto  $P_i$  é processado são construídos dois vetores  $P_{(i+k)} - P_i$  e  $P_{(i-k)} - P_i$ , e o ângulo  $\alpha_i$  entre os vetores é calculado. Se o menor ângulo formado entre os dois vetores for menor ou igual ao valor do limiar  $T_a$ , o ponto é considerado

como feição. Nesta etapa esses pontos não podem ser ainda considerados como sendo Picos ou Vales, mas apenas como feições ou pontos de interesse.

A seguir, na Figura 28, são mostrados os resultados da aplicação do algoritmo em imagens com diferentes posturas da mão. Os pontos reconhecidos como sendo feições são aqueles destacados na silhueta, em vermelho para os Picos e azul para os Vales. Os valores utilizados nos parâmetros de teste foram  $k = 21$  e  $T_a = 60^\circ$ . Como se esperava, vários pontos foram considerados como sendo feições da silhueta, pois os pontos das regiões vizinhas aos Picos e Vales possuem valores de k-curvatura e características semelhantes. A Figura 28.(b) mostra a detecção das feições com ligeiras mudanças de escala na silhueta da mão.

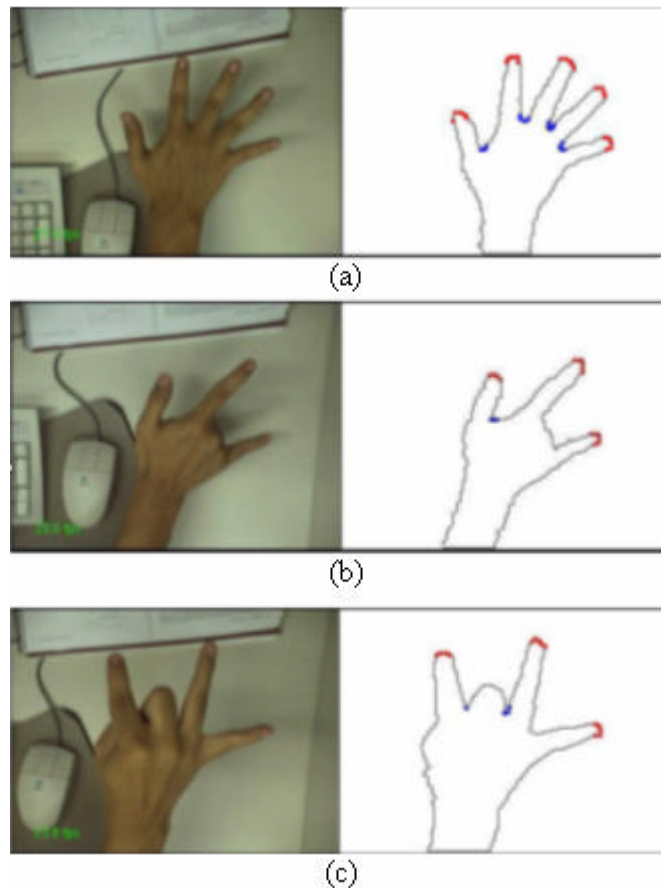


Figura 28 - Resultado da aplicação do algoritmo de detecção de feições à silhueta da mão. As feições são os pontos ressaltados na silhueta. (a) e (b) Mostram duas posturas da mão com os dedos estendidos. Os valores utilizados para o teste foram  $k = 21$  e  $T_a = 60^\circ$ . (c) Mesmo com pequenas mudanças de escala, aproximando-se a mão da câmera, e mantendo os mesmos parâmetros que em (a) e (b) pode-se ver a detecção das feições.

### 5.3. Detecção dos Dedos

Para que a detecção dos dedos seja um processo confiável e robusto, é importante ter como fonte feições que sempre possam ser encontradas na silhueta da mão. Os pontos detectados como sendo feições estão distribuídos em conjuntos isolados, os quais agrupam pontos com características semelhantes e representam um Pico ou um Vale na silhueta.

Na Figura 28 pode-se ver que todos os Picos da mão são detectados, mas não são detectados todos os Vales. A k-curvatura dos Vales é um valor que depende da abertura da junção entre os dedos e, visto que essa abertura pode ser alterada, os parâmetros para a detecção das feições nem sempre garantem sua detecção. Na Figura 28(b) pode-se ver um Vale que não é detectado porque a abertura da junção é maior do que o limiar para a k-curvatura. Já que os Vales nem sempre são detectados na silhueta, eles não são considerados fontes confiáveis para o reconhecimento. Desse modo, no processo de detecção dos dedos, os pontos que representam os Vales são descartados.

A detecção de um ponto que representa um dedo é o processo que analisa cada conjunto de pontos que representa os Picos e escolhe o ponto que possui o maior valor de k-curvatura dentro de cada conjunto. Assim, para cada conjunto de pontos que representa os Picos é detectado um ponto que representa um dedo. Na Figura 29 são mostrados os resultados desse processo aplicados a uma mão com três dedos estendidos. Agora existe apenas um ponto representando cada dedo.

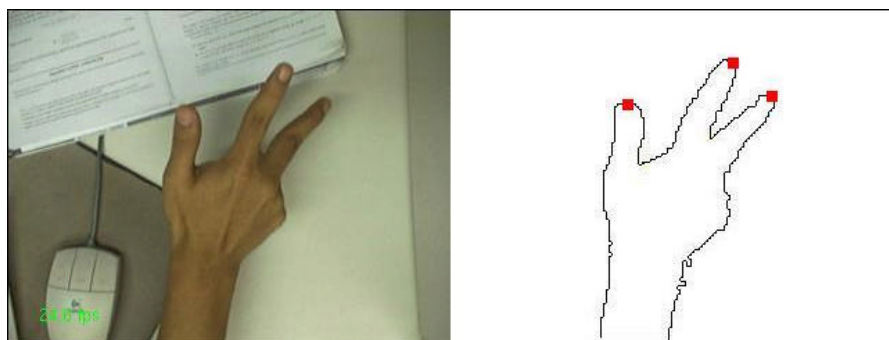


Figura 29 - Resultado do processo de detecção dos dedos da mão. Os pontos estão ressaltados na silhueta.

Além da posição 2D do ponto que representa o dedo, nesta etapa também é calculada a direção de cada um dos dedos. A direção é representada por uma linha que passa através do dedo e aproxima sua direção (Figura 30(a)). Cada

ponto  $P_i$  que representa um dedo teve como referência dois pontos, um anterior  $P_{(i-k)}$  e outro posterior  $P_{(i+k)}$ , para o cálculo da sua  $k$ -curvatura; esses pontos também serão úteis para o cálculo da direção do dedo que representa o ponto  $P_i$ .

Na construção da linha existem duas etapas: na primeira são construídos alguns pontos de referência ao longo do dedo e, na segunda, esses pontos são utilizados para construir a linha. Na primeira etapa, o trajeto  $[P_{(i-k)} P_i]$  e o trajeto  $[P_i P_{(i+k)}]$  da silhueta são divididos por  $n$  pontos igualmente espaçados. Depois, partindo do ponto  $P_i$ , são formados  $n$  pares de pontos juntando pontos a cada lado do ponto  $P_i$  (Figura 30(b)). Uma vez feito isso, os pontos de referência procurados são todos os pontos médios entre cada par de pontos. Na Figura 30(b) é mostrado um exemplo com 4 pares de pontos; os pontos de referência são marcados com uma “x”. Na segunda etapa, tendo esses pontos de referência “x” e o ponto  $P_i$ , é construída a linha de direção utilizando o método de Mínimos Quadrados descrito por Weisstein [43]. Assim, a direção do dedo é a linha que melhor se aproxima dos pontos de referência e que passa pelo ponto  $P_i$ .

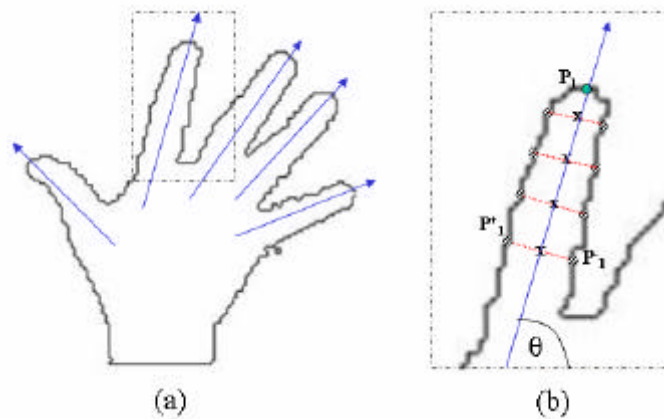


Figura 30 - Cálculo da direção dos dedos. (a) A direção aproximada de cada um dos dedos é mostrada como sendo uma linha. (b) A linha que representa a direção do dedo é a que passa pelo ponto  $P_i$  e que melhor aproxima os pontos de referência. Os pontos de referência “x” são os pontos médios de cada par de pontos que estão ligados por uma linha. O ângulo  $\theta$  é o valor que representa a orientação da linha de direção com respeito à horizontal.



Para representar numericamente a direção de cada dedo é utilizado o ângulo  $q$  que a linha faz com algum eixo principal da imagem. Neste caso o eixo principal escolhido foi a horizontal; assim, direções para a esquerda terão ângulos maiores de 90 graus e direções para a direita terão ângulos menores de 90 graus.

#### 5.4. Reconhecimento de Gestos

Antes de falar do processo de reconhecimento é importante definir um conjunto pré-determinado de gestos que serão reconhecidos. Conforme dito na primeira parte deste capítulo, o conjunto de gestos será representado por apenas 4 gestos da mão, os quais são mostrados na Figura 31. Esses gestos por sua vez estão divididos em três classes: *FECHADO*, *INDICE* e *ESTENDIDO*. O nome de cada uma das classes indica uma característica da mão com relação aos dedos. *FECHADO* é uma classe de gesto que não apresenta nenhum dedo estendido. Esse gesto, já considerando algumas aplicações, poderia ser utilizado para indicar algum tipo de inatividade. *INDICE* é uma classe de gesto que apresenta um ou dois dedos da mão estendidos. Esse gesto é utilizado principalmente para interagir com as aplicações. *INDICE* engloba dois gestos parecidos, os quais são chamados de *INDICE\_1* e *INDICE\_2*, cuja única diferença é a falta do dedo polegar em *INDICE\_1*. *ESTENDIDO* é uma classe de gesto que apresenta mais do que dois dedos estendidos.

As três classes de gestos são bem diferentes entre si no que respeita ao número de dedos e forma. Esta característica faz com que o processo de reconhecimento de cada classe seja mais eficiente e mais simples ao mesmo tempo. Mesmo nos gestos de mesma classe (*INDICE\_1* e *INDICE\_2*) existem diferenças grandes para poder reconhecer cada gesto em particular.

O processo de reconhecimento dos gestos é um processo de classificação baseado em feições locais situadas ao longo da silhueta da mão. Essas feições são os pontos que representam os dedos e que foram calculados na etapa anterior.

Para detectar um gesto é feita uma operação de comparação entre o número de dedos detectados e o número de dedos presentes em cada um dos gestos predefinidos para o reconhecimento. Nesta etapa se espera que os únicos gestos utilizados sejam os 4 gestos previamente definidos. Dentre os 4 gestos reconhecíveis existe um gesto — *FECHADO* — que apenas possui

informação gestual e não possui parâmetros dos dedos e valores que possam ser aproveitados.

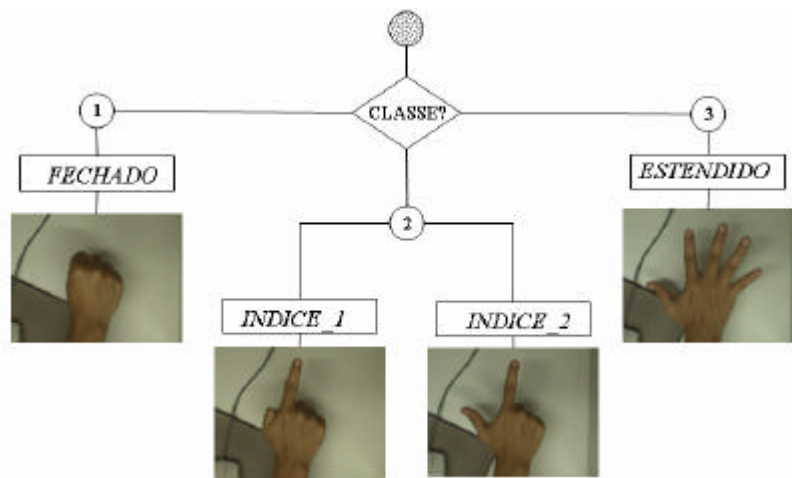


Figura 31 - Conjunto predefinido de gestos que são reconhecidos. Os 4 gestos são divididos em três classes: FECHADO, INDICE e ESTENDIDO. FECHADO é uma classe de gesto que não apresenta nenhum dedo estendido. INDICE representa uma classe com dois tipos de gestos: INDICE\_1 (dedo índice estendido) e INDICE\_2 (dedo índice e polegar estendidos). ESTENDIDO é uma classe de gesto que apresenta mais de dois dedos na mão estendidos.

Quando não são detectados dedos na mão, o gesto inferido é o *FECHADO*. Quando são detectados mais do que dois dedos na mão, o gesto inferido é o *ESTENDIDO*. Se forem detectados um ou dois dedos, então, dependendo do número dos dedos, é inferido um gesto da classe *INDICE*. Este processo de inferência de gestos baseado no número de dedos pode também ser representado como sendo o reconhecimento de *tokens* feito por um Autômato Finito Determinista (Figura 32). Nesse autômato existe um único símbolo no vocabulário, *d*, que significa “*dedo*”. Há 4 estados diferentes, sendo três deles, *q1*, *q2* e *q3*, estados finais ou de reconhecimento. O autômato só é testado quando existem cadeias de símbolos *d* que podem ser reconhecidos. No caso da mão *FECHADA* o autômato não é testado. Os estados de reconhecimento *q1*, *q2* e *q3* correspondem aos gestos *INDICE\_1*, *INDICE\_2* e *ESTENDIDO*, respectivamente.

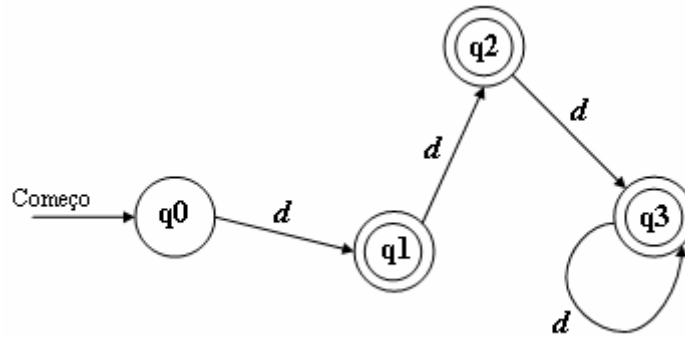


Figura 32 - Autômato Finito Determinista que representa o processo de inferência dos gestos com base no número de dedos “d” detectados na silhueta da mão. O único símbolo do vocabulário é “d”, que significa “dedo”. Os estados de reconhecimento **q1**, **q2** e **q3** representam os gestos INDICE\_1, INDICE\_2 e ESTENDIDO, respectivamente.

Ao contrário dos gestos *FECHADO* e *ESTENDIDO*, nos gestos *INDICE* é necessário descobrir qual dos dedos é o índice. No caso de ser detectado apenas um dedo, este é assumido como sendo o índice. Quando dois dedos são detectados, o dedo índice é aquele que está à direita do outro (isto se aplica à mão direita). Já que os pontos da silhueta são armazenados na ordem anti-horária, convém mencionarmos o conceito de Distância na Silhueta  $DNS(P_i, P_j)$  como sendo o número de pontos da silhueta entre o ponto  $P_i$  e o ponto  $P_j$ . Utilizando uma heurística simples podemos dizer que se  $DNS(P_i, P_j) < N/2$  então  $P_i$  é o dedo índice e está à direita de  $P_j$ ; se não,  $P_j$  é o dedo índice e está à direita de  $P_i$ .  $N$  é o número total de pontos da silhueta.

Como os gestos são inferidos com base em feições locais ao longo da silhueta, diferentes gestos da mão com o mesmo número de dedos podem ser inferidos como sendo um mesmo gesto. Se um gesto diferente aos gestos predefinidos é mostrado, a resposta da inferência dos gestos será aquele gesto predefinido que mais se assemelhe, em número de dedos, com o gesto sendo analisado.

## 5.5. Resultados Parciais

Nesta seção são apresentados os resultados de alguns testes feitos com o reconhecimento dos gestos da mão. Já que esse reconhecimento se baseia no número de dedos reconhecidos, os primeiros testes foram direcionados para

avaliar o resultado do número de dedos reconhecidos na mão e os demais visaram avaliar o reconhecimento dos gestos predefinidos.

Para avaliar o reconhecimento do número de dedos foi considerado um intervalo de tempo, de aproximadamente um minuto, contendo um determinado número de quadros. Durante esse tempo, dependendo do número de dedos testados, a mão foi mantida com um número específico de dedos. Ao final desta etapa foi contabilizado o número de acertos e erros no reconhecimento dos dedos. Com base nesses valores foram calculadas as porcentagens de acerto e erro para cada número de dedos presente na mão. A Tabela 1 mostra os resultados dos testes realizados para avaliar o reconhecimento do número de dedos presente na mão.

<b>Número de Dedos</b>	<b>Total de Quadros</b>	<b>Total de Acertos</b>	<b>Total de Erros</b>	<b>Porcentagem de Acerto</b>	<b>Porcentagem de Erro</b>
0	1136	1136	0	100.0 %	0.0 %
1	1100	1096	4	99.6 %	0.4 %
2	1234	1212	22	98.2 %	1.8 %
3	1350	1335	15	98.9 %	1.1 %
4	1057	1020	37	96.5 %	3.5 %
5	1094	1010	84	92.4 %	7.6 %

Tabela 1 - Resultado dos testes feitos para a avaliação do reconhecimento do número de dedos presente na mão. Para cada número de dedos é apresentada uma estatística indicando o número e a porcentagem de acertos e erros ao longo do teste.

Para a avaliação do reconhecimento dos gestos predefinidos, da mesma forma que no teste anterior, foi levado em conta um intervalo de tempo durante o qual o gesto da mão foi mantido. Ao final, foi contabilizado o número de acertos e erros no reconhecimento de cada um dos gestos. A Tabela 2 mostra os resultados da avaliação para o reconhecimento dos gestos predefinidos.

<b>Gestos</b>	<b>Total de Quadros</b>	<b>Total de Acertos</b>	<b>Total de Erros</b>	<b>Porcentagem de Acerto</b>	<b>Porcentagem de Erro</b>
<i>FECHADO</i>	1305	1305	0	100.0 %	0.0 %
<i>INDICE_1</i>	1261	1255	6	99.5%	0.5 %
<i>INDICE_2</i>	1054	1044	10	99.04 %	0.96 %
<i>ESTENDIDO</i>	1125	1114	11	99.02 %	0.98 %

Tabela 2 - Resultado dos testes feitos para a avaliação do reconhecimento dos gestos predefinidos. Para cada um dos 4 gestos é apresentada uma estatística indicando o número e a porcentagem de acertos e erros ao longo do teste.

## **5.6. Considerações Finais do Capítulo**

Para a etapa do reconhecimento foi considerado apenas um conjunto de 4 gestos. Esses gestos procuram ser gestos básicos que ajudem no processo de reconhecimento e permitam mostrar que, a partir de um conjunto pequeno, podem ser construídas diferentes aplicações.

As únicas feições consideradas foram os Picos e Vales. A silhueta possui mais informações relevantes que também podem ser consideradas, mas requerem abordagens mais específicas para procurar outras feições. Características globais podem ser utilizadas como base para construir modelos de reconhecimento mais sofisticados baseados na silhueta.

Neste trabalho procuramos caracterizar os gestos e posições com apenas as feições de Picos e Vales. Apesar de serem só duas, muitos gestos podem ser reconhecidos de forma estável.