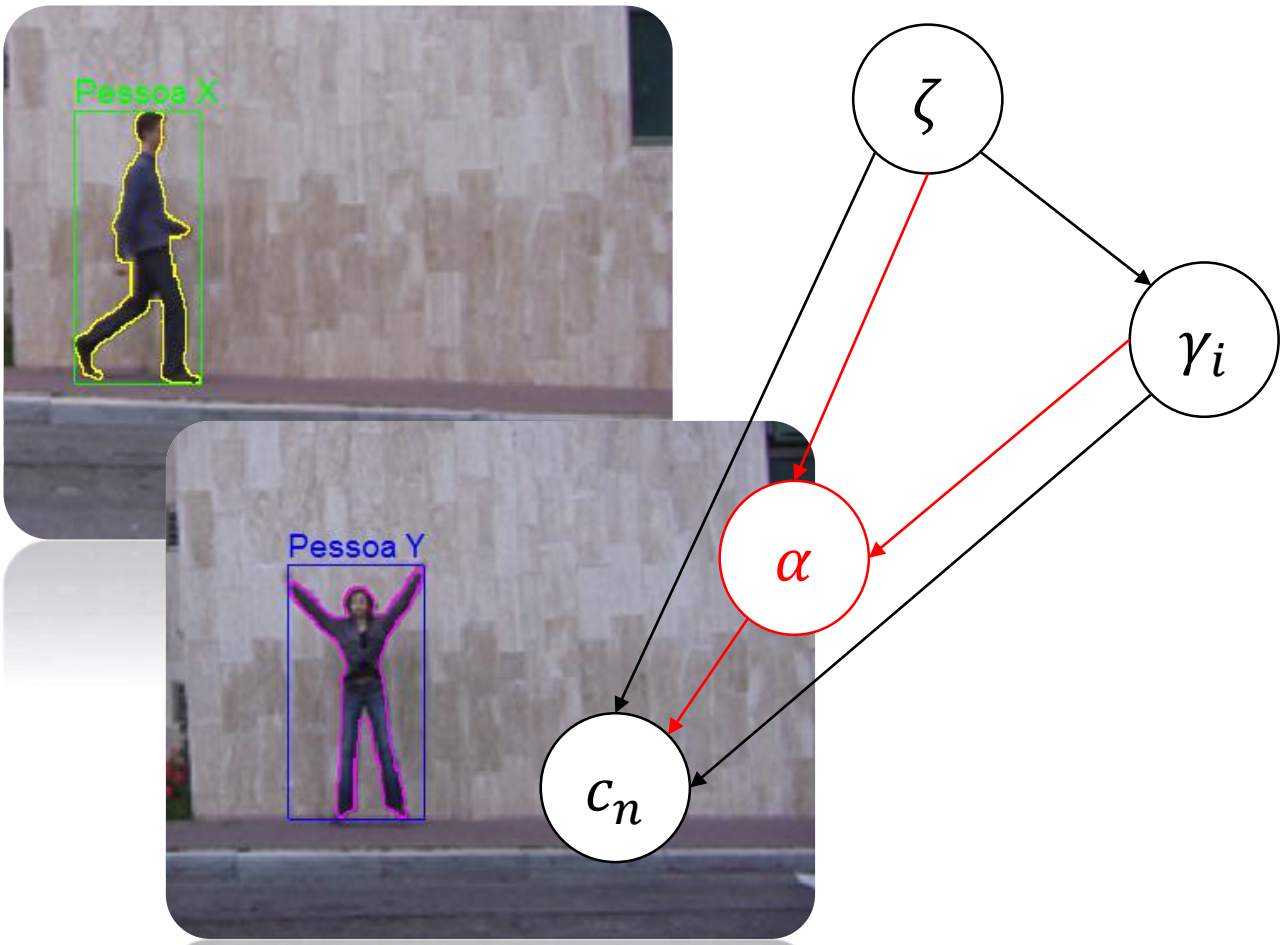


José António Monteiro de Sousa

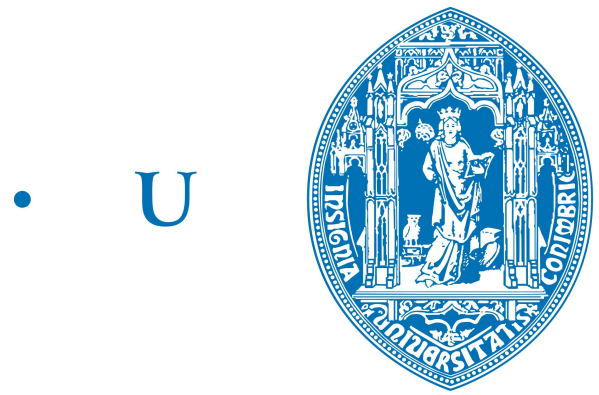
Sistema de Identificação de Pessoas com base no seu Movimento

Dissertação de Mestrado em Engenharia Electrotécnica e de Computadores

Julho de 2013

· U C ·

Universidade de Coimbra



Sistema de Identificação de Pessoas com Base no seu Movimento

José Sousa

Coimbra, Julho de 2013

Dissertação de Mestrado Submetida ao Dep. de Engenharia Electrotécnica e Computadores

Faculdade de Ciências e Tecnologia
Universidade de Coimbra

Júri:

Presidente	: Doutor Vítor Manuel Mendes da Silva
Vogal	: Doutor Paulo Jorge Carvalho Menezes
Orientador	: Doutor Jorge Manuel Miranda Dias
Co-orientador	: Eng. Luís Carlos Santos

Resumo

Nesta tese é proposto um modelo probabilístico para a Análise de Movimento de Laban Visual, baseado em silhuetas do corpo humano, como uma solução estendida, dentro da área de investigação que estuda identificação de pessoas. Os modelos para Laban presentes na literatura, são sobretudo baseados em características adquiridas de tecnologias muito precisas, sendo que o método aqui proposto é sugerido como uma aproximação baseada em sistemas de visão gerais, generalizando o modelo de Laban para características visuais. Dada a potencial dimensão da silhueta, é sugerida uma representação alternativa, baseada na Análise das Componentes Principais e na Transformada Generalizada de Fourier. Os descritores simbólicos de Laban são aprendidos com base em conjuntos de dados de treino, e são posteriormente usados para caracterizar sequências desconhecidas, usando um classificador Bayesiano. O método proposto foi integrado com sucesso num sistema biométrico não invasivo de identificação de pessoas, criando as bases para o desenvolvimento de uma plataforma de vídeo vigilância inteligente. Contudo, a integração directa do modelo Visual de Laban com o sistema de identificação existente, provou não ser uma solução adequada. A análise de resultados preliminares, levou a que fosse proposta uma nova e modificada solução, que foi capaz de lidar com os problemas resultantes da generalização da Análise de Movimento Laban ao corpo como uma entidade única, em detrimento da parametrização em partes do corpo específicas. Os resultados experimentais demonstram um modelo de classificação preciso, em que as características mostram ser discriminantes relativamente aos diferentes actores, quando executado diversas acções. O sistema experimental proposto mostra capacidade de analisar movimento bem como identificar pessoas, resultados validados em duas bases de dados experimentais públicas e reconhecidas na área.

Abstract

In this thesis, a probabilistic model for *Visual* Laban Movement Analysis, based on human body silhouettes, is proposed as an extended solution within the problematic of person identification. Laban models presented in literature are mostly based on features acquired from precise tracking technologies, whereas our method is suggested as a vision-based approach, generalizing the Laban model to visual-cues. Given the silhouette's dimension, we suggest an alternative representation based on Principal Component Analysis and the generalized Fast Fourier Transform. Laban symbolic descriptors are learned based on sets of training data, which are posteriorly used to characterize unknown sequences applying a Bayesian classifier. The proposed method is successfully integrated with a person identification framework, creating the grounds for developing an intelligent video-surveillance system. The direct integration of the visual Laban model with the existent identification framework, proved to be a lesser solution. Therefore, we propose a new, modified version, which was able to cope with the issues inherent to the proposed Laban generalization addressing the body as a single entity, rather than parametrized into specific body parts. Results demonstrate an accurate classification model, in which features show to be discriminant to different performers, when performing the multiple actions. The experimental set-up shows capability for motion analysis and person identification, validated in two public and acknowledgeable datasets in the area.

Agradecimentos

Ao Professor Doutor Jorge Manuel Miranda Dias expresso, o meu profundo agradecimento, pela orientação e apoio que elevaram os meus conhecimentos científicos e estimularam o meu desejo de querer sempre saber mais e a vontade constante de querer fazer melhor.

Ao Eng. Luís Santos pela sua permanente disponibilidade, a maneira como me ajudou a prosseguir na dissertação com a sua combinação perfeita e única de críticas, sugestões e incentivos.

A toda a minha família e amigos por todo o apoio tanto ao longo do curso como da minha vida. Espero de alguma forma retribuir todo o carinho e apoio que sempre me habituaram.

A todos vocês, dedico este trabalho. O meu eterno agradecimento.

Indice

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

As câmaras de vídeo são cada vez mais prevalentes na sociedade, tanto em espaços públicos como privados. Simultaneamente, a qualidade da vídeo vigilância continua a aumentar. Isto é verdade, especialmente para a tecnologia de sistemas de vídeo vigilância inteligentes, que consegue reconhecer ou seguir objectos, bem como identificar caras, gestos ou padrões de comportamento. [Held et al., 2012].

Uma das área tecnológicas bastante activa da nossa economia, é sem dúvida aquela que se dedica aos sistemas de segurança, sendo que uma das sub-áreas de maior visibilidade e investigação é sem duvida a vídeo vigilância. Estes sistemas são de tal forma explorados, que naturalmente começou a surgir a necessidade de desenvolver sistemas inteligentes para o efeito. Esta característica permite ao sistema a capacidade de dar uma resposta mais rápida ou mesmo de prever situações potencialmente alarmantes. Alguns exemplos do uso deste sistemas são os centros comerciais, parques de estacionamento, escolas, bancos, bem como locais privados ou restritos onde a circulação de pessoas é limitada.

Na área de processamento de vídeo e imagem digital existe actualmente uma grande variedade de algoritmos de detecção de movimento e de objectos. Detecção automática de pessoas e interpretação dos seus movimentos em sequências de vídeos é uma delas. Sendo um tema tão actual, este problema é amplamente abordado não só na área de vídeo vigilância, mas também, por exemplo, em sistemas inteligentes, consolas, etc. A complexidade do problema passa muito pela dificuldade de modelar uma pessoa e as suas actividades, dada a variabilidade presente nos vários movimentos de um corpo humano. Uma pessoa pode apresentar um enorme conjunto de aparências físicas, poses, movimentos e acções entre outras pessoas ou objectos. Actualmente existem muitos sistemas que abordam e tentam mitigar este tipo de problemas.

O estado de arte sobre detecção e seguimento de pessoas tem já muitas soluções com elevado sucesso nesta matéria. Nos últimos anos, a implementação de sistemas automáticos com a capacidade de identificar pessoas tornou-se numa área em intensa pesquisa. Contudo detecção de actividades (que não a actividade de andar/correr) com objectivo de identificação discriminativa de pessoas é uma área que apenas recentemente começou a ser explorada, muito devido a toda a sua complexidade. Define-se este problema como identificação de pessoas com base nas características do seu movimento. A sua aplicação mais evidente será a integração em sistemas de vídeo vigilância, onde um sistema de sensores, geralmente

câmaras, adquire imagens de diferentes pessoas com o objectivo de as identificar e, caso necessário, despoletar respostas adequadas.

Actualmente já existem diversas aplicações, no âmbito da biometria, capazes de fazer este tipo de reconhecimento automático de pessoas, que na sua maioria são sistemas de identificação biométrica invasivos. Estes sistemas incidem principalmente em reconhecimento facial ou identificação através da impressão digital ou íris. Apesar destas aplicações serem extremamente bem sucedidas, obrigam a um comportamento cooperativo, ou seja, a uma interação normalmente voluntária do utilizador com o próprio sistema. Por isso o reconhecimento de pessoas através das suas actividades físicas e motoras é uma área de pesquisa em biometria não invasiva, tópico este, onde também se insere o trabalho proposto nesta tese.

1.1 Identificação do Problema

Este trabalho pretende assim dar continuidade ao trabalho realizado por [Santos and Dias, b],no qual dois grandes problemas são identificados.

- O primeiro grande problema versa sobre os modelos de Análise ao Movimento Laban que são na sua maioria, baseados na cinemática ou características dinâmicas do movimento, ao passo que características visuais clássicas ainda não foram exploradas nesse sentido. Define-se então o seguinte problema: dada uma imagem, ou sequência de imagens, extrair um conjunto de características quem seja discriminantes relativamente aos vários parâmetros de Laban. Estas características deverão depois ser usadas como evidências num classificador, com o objectivo de se obter uma análise simbólica/semântica do movimento observado, com base em parâmetros Laban.

- O segundo grande problema insere-se na área de investigação em sistemas de identificação biométricos. Sistemas de identificação de pessoas com base no seu movimento são ainda bastante restritos à actividade de andar. Neste sentido, podemos formular o seguinte problema: com base na caracterização simbólica do movimento da solução do problema anterior, gerar um conjunto de assinaturas de movimento que realçem as características com que cada pessoa se movimenta. Essas assinaturas devem ser então usadas num algoritmo de classificação, de modo a que se consigam identificar diferentes pessoas com base no seu movimento.

A solução para estes dois problemas, permitirá estabelecer uma base sólida para o desenvolvimento de sistemas de vídeo vigilância inteligentes, com capacidade para análise e reconhecimento de actividades, bem como capacidade para identificar diferentes pessoas pelas suas características motoras.

1.2 Solução Proposta

Propõe-se então uma solução através da abordagem ilustrada na Figura 1.1. O processo passa por adquirir um conjunto de sequências de imagens a partir de um conjunto de câmaras, onde as silhuetas são segmentadas. Dada a sua dimensionalidade propõe-se uma

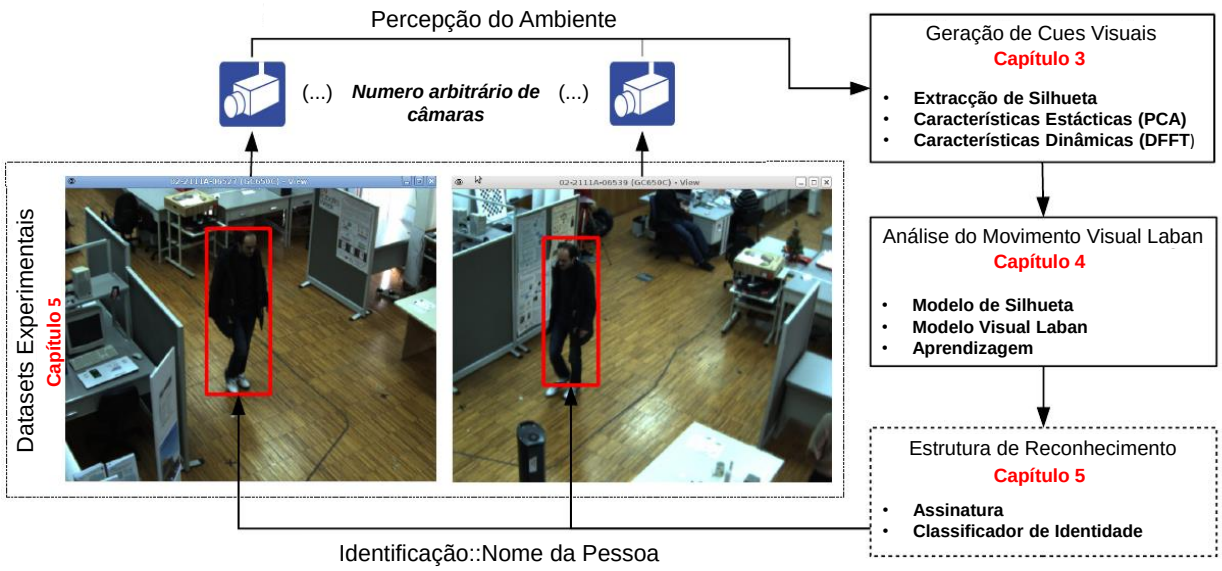


Figura 1.1: Estrutura do trabalho proposto

representação alternativa, que combina duas técnicas de processamento de sinal: Análise de Componentes Principais (PCA) e Transformada Rápida de Fourier (FFT). Posteriormente é aplicada uma estratégia de aprendizagem supervisionada para associar as características das silhuetas geradas para um conjunto de dados treinados, que são manualmente classificados com as características Laban dominantes. Ao aprender o modelo visual de análise de movimento Laban usa-se um classificador baseado nas Redes Dinâmicas Bayesianas, que é aplicada a sequências de movimentos autonomamente analisados, usando descritores simbólicos Laban. Resolvendo os desafios anteriormente mencionados identificamos as seguintes potenciais contribuições:

- Um modelo de Análise de Movimento Laban baseado em características visuais.
- Apresentação de uma assinatura Laban adaptada para identificação de pessoas tendo por base as características das pessoas inerentes à forma de se movimentarem.

1.3 Trabalhos Relacionados

Apresentam-se de seguida alguns dos trabalhos mais recentes e significativos na área de identificação de pessoas com base no seu movimento, bem como uma enumeração das principais técnicas aplicadas em cada um deles.

[Iosifidis et al., 2012] apresenta um método invariante à escala de identificação de pessoas, explorando a informação obtida através de um sistema multi-câmaras e incorporando algumas actividades no processo de identificação de pessoas. Um sistema multi-câmaras é usado para captar o corpo humano em diferentes ângulos. *"Fuzzy Vector Quantization"* e *"Linear Discriminant Analyis"* são técnicas exploradas para criar representações detalhadas da actividade. São obtidos resultados da identificação de pessoas, reconhecimento de actividade e ângulo de visão para cada câmara independentemente.

[Yam et al., 2002] desenvolveu um modelo invariante de movimentos para as acções *andar* e *correr*, com recurso a técnicas usadas em visão por computador para capturar o movimento da perna enquanto o sujeito anda ou corre. Para isso, recorre a uma análise de Fourier do movimento para obter características importantes para o reconhecimento, e também como análise entre a relação do *andar* e *correr*. Foi concluído que a relação entre *andar* e *correr* existe, podendo ser descrito de uma forma genérica e que a relação entre ambas pode estar relacionada com a modulação de fase. Correr é também uma característica única e pessoal , já que esta é uma das características mais genuínas de cada sujeito.

[Kouno et al., 2012] trata de um sistema de tarefas de identificação de pessoas através de uma câmara no topo do sujeito. Com esta acção, reduziram a restrição da localização da instalação da câmara e resolveram problemas de oclusão de imagem. Utilizaram também informações de profundidade para a identificação. São aplicadas quatro características para o método de identificação: dimensão, altura, tamanho estimados dos corpos e histograma de profundidade. Apesar de elevada eficácia nos resultados obtidos, obtiveram variações elevadas de valores da distância obtidos pela câmara (Kinect), estimando assim valores errados das características. Para resolverem esse problema, aplicaram um processo de normalização, para obter resultados mais satisfatórios (melhoria na ordem dos 20%). Assim é importante o processo de normalização. Outra lacuna prende-se no facto de as dimensões e tamanho do corpo serem sensíveis ao tipo de roupa usado. Este problema é um ponto de interesse futuro para o seu trabalho, já que a informação sobre a roupa tem um papel importante na identificação de pessoas como informações contextuais, e é nessa característica contextual que pretendem aplicar ao método já tratado.

[Hamdoun et al., 2008] apresentam e avaliam um programa de re-identificação de pessoas para um sistema de vigilância com multi-câmaras. O método é baseado em correspondências de assinaturas, baseadas em descritores de pontos de interesse, obtidos em pequenas sequências de vídeo. Uma das originalidades do método é o facto de acumular pontos de interesse em algumas imagens espaçadas no tempo durante o processo de acompanhamentos do sujeito em cada câmara, com o objectivo de captar aparências variáveis.

[Iwashita and Kurazume, 2009] propõem um método híbrido de identificação de pessoas usando biometrias fisiológicas e comportamentais. Esta técnica combina a forma do corpo (fisiologia) e a marcha do sujeito (comportamento) e trata todo o volume espaço-temporal esculpido de uma pessoas que anda. A esta biometria extraem também biometrias individuais únicas, criando imagens médias do volume do espaço temporal e formando novos volumes

de espaços temporais de diferentes imagens, que são criadas ao subtrair uma imagem média de imagens originais. Mostrou que combinando momentos invariantes com classificador *"Support Vector Machine"* obtém-se resultados ligeiramente melhores do que apenas baseado apenas na análise de marcha e na classificação "K-Nearest neighbour".

[Yu et al., 2006] propõem um sistema para enfrentar o problema de ainda não existirem métodos padrão para comparar a performance de diferentes algoritmos de reconhecimento da maneira de andar. Este quadro consiste numa base de dados com marchas de 124 sujeitos onde em cada marcha de cada sujeito foram captadas 11 vistas diferentes. São separadas na base de dados, o ângulo, roupa e mudanças de condições de transporte. A base de dados usada é ambiciosa dado o seu tamanho. A base de dados pode ser usada para reconhecimento de sujeitos pela maneira de andar, reconstrução do corpo humano, análise do movimento do sujeito, etc. De futuro pretende-se explorar esta técnica em ambientes exteriores, com outro tipo de sensores e em tempos diferentes com uma base de dados ainda mais ampla. Têm também como objectivo uma avaliação sistemática para o reconhecimento da marcha de um sujeito, como *"Face Vendor Recognition Test"* no reconhecimento facial.

[Han and Bhanu, 2006] propõem uma nova representação espaço-temporal de marcha, *"Gait Energy Image"*, para caracterizar propriedades do andar humano para reconhecimentos individuais, através da sua marcha. GEI representa sequências de movimento humano numa simples imagem enquanto preserva informação temporal. Ju Han e Bhanu propõem também um reconhecimento humano combinando modelos sintéticos com estatísticas de características de marcha reais. Para ultrapassar limitações de modelos de treino, propõe um simples modelo para simular distorção de modelos sintéticos e uma fusão de características de marcha para reconhecimento humano através da marcha do sujeito. GEI tem uma representação eficiente da marcha e a aproximação do reconhecimento proposto atinge boas performances. Algumas limitações persistem nesta técnica como a roupa ou sapatos usados, ou até o próprio contexto envolvente. Condições físicas como pequenas lesões motoras podem também mudar a maneira de andar do sujeito. Uma grande variação de marcha de uma só pessoa em diferentes condições reduz a qualidade de uma marcha como biometria, podendo não ser única como detecção por íris ou impressão digital, mas a característica de marcha do indivíduo continua a ser muito usada em vigilância visual.

[Santos and Dias, b] apresentam um modelo baseado em trajectórias de partes do corpo, onde se desenvolvem várias técnicas para geração de assinaturas de movimento, com o objectivo de serem usadas para discriminar diferentes identidades. Trajectórias geométricas e parâmetros dinâmicos são associados a diferentes símbolos através de uma aprendizagem modular baseada em *Gaussian Mixture Model*. O sistema apresentado é faseado em duas partes. A primeira usa classificadores Bayesianos para estimar conjuntos de parâmetros simbólicos, que depois são associados num algoritmo capaz de gerar assinaturas contendo características únicas sobre a maneira como cada pessoa executa movimento. A segunda usa as ténicas de assinatura propostas para, com um classificador, conseguir identificar pessoas.

1.4 Estrutura da Tese

A estrutura deste trabalho será a seguinte:

- *CAPÍTULO I*: Este primeiro capítulo apresenta uma introdução e estado de arte relacionado com o trabalho apresentado. Enuncia também os objectivos para o desenvolvimento de todo o trabalho.
- *CAPÍTULO II*: Este é um capítulo introdutório ao tema Laban. Irá falar-se dos tipos de notações existentes e do tipo de notação que será incluído neste trabalho.
- *CAPÍTULO III*: Este capítulo trata os tipos de processamento de imagem utilizados e da extracção de características relevantes no sentido de serem incluídas nos capítulos seguintes.
- *CAPÍTULO IV*: Neste capítulo é criado um modelo Bayesiano de Laban para caracterizar vários tipos de movimentos executados por diferentes pessoas.
- *CAPÍTULO V*: Neste capítulo é incorporado um sistema de geração de assinaturas e um modelo de assinaturas com vista a obtenção de uma identificação correcta de pessoas.
- *CAPÍTULO VI*: Este capítulo sumariza os objectivos principais do trabalho, sendo discutidos os resultados e apresentando sugestões para trabalho futuro.

1.5 Acrónimos

Nesta secção apresentam-se os acrónimos usados durante esta tese e os seus significados.

Tabela 1.1: Lista de acrónimos e seus significados.

Acrónimo	Significado
PCA	Análise de Componentes Principais
LMA	Análise de Movimento Laban
FFT	Transformada Rápida de Fourier
DFT	Transformada de Fourier Discreta
MAP	Maximum a Posteriori

Nota Importante: Para manter uma coerência semântica adequada, preservaram-se os termos em inglês dos estados das componentes e qualidades do espaço Laban. Mantém-se também alguns termos em inglês cuja tradução em português possa não ser considerada precisa.

Capítulo 2

Análise do Movimento Laban e Notação

O movimento do corpo humano pode ser definido como o processo que consiste em mover uma ou mais partes do corpo, arbitrariamente ou não, para uma localização específica, ao longo de uma trajectória no espaço. Em muitas situações, as pessoas são capazes de, com a observação do movimento que outra pessoa executa, o caracterizar, ou fazer interpretações acerca das propriedades geométricas e dinâmicas. Sobre as mesmas, somos capazes de inferir comportamentos, estados de espírito ou simplesmente que identificar o tipo de gesto executado, só para dar alguns exemplos. Este tipo de racionalização, é geralmente aproximada pela comunidade científica através de algoritmos de classificação, que se enquadram na área do reconhecimento de padrões. Habitualmente, esta tarefa de classificação (racionalização) sobre o movimento é suportada por evidências adicionais que não podem ser descritas apenas com base em informação cinemática. O movimento corporal é rico também em informação de cariz psicológico: expressividade do movimento. Esta caracteriza o movimento, e pode surgir de forma consciente ou inconsciente em reacções instintivas. Dá-se um exemplo bastante simples, onde na prática de exercício físico, uma pessoa executa os seus movimentos de uma forma natural e inconsciente (Figura 2.1a). Pelo contrário, em actividades de elevado nível de concentração no movimento, como um espectáculo de dança (Figura 2.1b, a expressividade é muitas vezes pré-definida pelos requisitos técnicos, sendo fundamental para o artista conseguir transmitir certas emoções através do seu movimento.

[Kendon, 2004] destaca alguns aspectos da acção corporal e trata-os como *gestos*. No seu trabalho, afirma que os *gestos* têm uma maneira própria de acção e algum significado expressivo. O sujeito é imaginado a exercer controlo de, pelo menos, alguns graus de liberdade sobre o movimento e o que o expressa. Estas definições permitem a distinção entre *gestos* e movimentos, no que respeita ao movimento se caracterizar por ser nervoso, habitual ou involuntário, e às acções necessárias durante a sua interacção. Apesar de dominada por comunicação verbal, a interacção social entre pessoas é rica em comunicação não verbal, em que sinais expressivos são transmitidos através do movimento. Olhemos para a política, onde o orador detém geralmente uma grande capacidade expressiva, usando as mãos e sinais subtis, como acenar com a cabeça, para vincar certos pontos chave do seu discurso.

Para o desenvolvimento de um modelo matemático que represente actividade corporal, é necessária uma definição teórica do mesmo. O movimento corporal vem sendo estudado à séculos, sendo que é geralmente descrito teoricamente usando algum tipo de notação. De

seguida faz-me uma introdução a algumas das notações de movimento mais populares.

2.1 Tipos de notações

Desde o século XVII que os sistemas de notação são conhecidos. Pierre Beauchamp e Raoul Auger Feuillet começaram em 1700 um programa de publicação de danças simbolizadas [Little and Marsh, 1992]. Os elementos básicos da notação de Beauchamp-Feuillet são o trato, símbolos para a posição dos pés, símbolos para o passo, sinais para acções, rotações e ritmo.

A notação do movimento Benesh foi inventada no final dos anos de 1940 por Joan e Rudolf Benesh para documentar qualquer tipo de dança ou movimento humano [Benesh and Benesh, 1983]. Usa uma pauta de cinco linhas que é lida da esquerda para a direita, onde barras verticais marcam a passagem do tempo, assemelhando-se assim a uma pauta musical. Toda a informação sobre a posição do corpo e membros é mostrada dentro da pauta. Linhas de movimento traçam o caminho feito pelas extremidades. Linhas de locomoção ligam as posições dos pés, mostrando se o executante dá passos, saltos ou desliza de uma posição para a outra. O ritmo e a qualidade do movimento é mostrada acima da pauta.

A notação do movimento Ehkol-wachman foi desenvolvidas pelo coreógrafo Noa Eshkol e pelo arquitecto Abraham Wachman [Eshkol and Wachmann, 1958]. Era usada para analisar o comportamento animal assim como a dança. O sistema utiliza uma representação de uma figura de linhas para descrever a orientação dos diferentes membros do corpo. As posições são expressas num sistema de coordenadas esféricas usando uma componente horizontal e vertical. É frequentemente discretizada em unidades de segmentos de 45º que são numerados de 0 a 7. As coordenadas horizontal e vertical dadas pela espera são escritas uma acima

(a) Jogging.

(b) Ballet.

Figura 2.1: Exemplos de actividades de expressividades de origens diferentes: (a) natural, e (b) controlado.

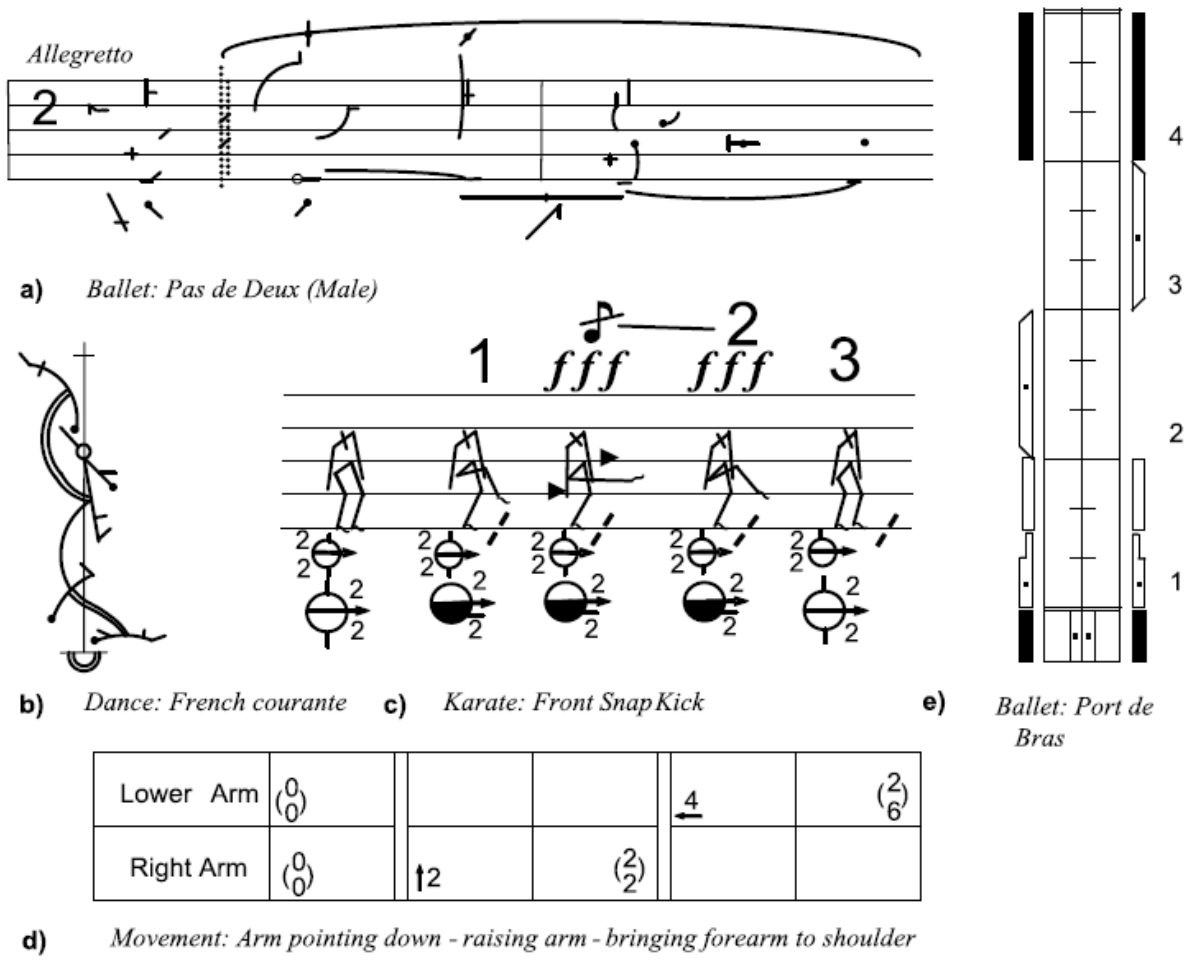


Figura 2.2: Ilustração de diferentes tipos de notações

da outra. As unidades do tempo são representadas em grelhas da esquerda para a direita, e os membros escritos numa linha diferente, a partir do topo e até ao fundo. A notação do movimento Ehkol-wachman representa um bom descritor para posições espaciais e a cadeia cinemática não está limitada ao corpo humano de uma forma conjunta. O conteúdo expressivo dos movimentos não são descritos.

Contudo, as notações apresentadas são de foco geométrico, isto é, colocam a sua descrição nas trajectórias e forma que o corpo tem de tomar durante uma dada execução. Neste sentido, surge a Análise de Movimento Laban, com uma notação semanticamente rica, que permite além da caracterização geométrica, a análise de características expressivas do movimento.

Notação de Laban, baseada na **Análise de Movimento Laban**, é um sistema de gravação de movimentos originada por Rudolf Laban, em 1920. Através deste método, todas as formas de movimento, desde a mais simples à mais complexa podem ser escritas com precisão. Este tipo de sistema pode ser aplicado em qualquer campo onde seja necessário gravar todo o deslocamento realizado pelo corpo humano. Áreas como a dança, antropologia, deporto e fisioterapia são alguns exemplos onde este tipo de sistema pode ser usado para caracterizar o movimento. Funcionando como que multi-ferramenta, a Notação de Laban

fornece uma maneira significativa de gravar movimento no papel para referência futura, uma análise de movimento de forma familiar e uma terminologia cuidadosamente seleccionada, tornando-se universalmente aplicável. Proporciona um entendimento universal do movimento servindo assim como uma linguagem comum para que todos os trabalhadores nas diferentes áreas e diferentes países possam comunicar. O sistema é assim como uma *Pedra de Roseta*(pedra egípcia), onde o conteúdo da cinética de todas as formas de movimento e estilos podem ser compreendidos. Elementos comuns podem ser distinguidos e diferentemente anotados. A sua simbologia não verbal, ainda que associada a uma semântica extremamente rica, mitiga qualquer tipo de barreiras linguísticas favorecendo assim o intercâmbio.

2.2 Análise de Movimento Laban

Análise de movimento Laban é um sistema de linguagem para compreender, observar, descrever e analisar todas as formas de movimento. Imaginada por Rudolf Laban, a Análise do Movimento Laban oferece um vocabulário para descrever características física e estrutural do movimento do corpo, o uso do espaço e da dinâmica, e os aspectos expressivos e qualitativos do movimento. Ao contrário das outras notações, Análise de Movimento Laban inclui uma análise aprofundada da dinâmica do movimento. Existem assim três formas essenciais de análise e descrição:

- *Descrição de Motivo*
- *Descrição Esforço-Forma*
- *Descrição Estrutural*

Motivo é forma mais simples da descrição. Fornece uma definição mais geral sobre o tema ou a característica mais saliente de um movimento. Também aponta a motivação do movimentos, as suas ideias, objectivos ou intenções. A *Descrição Motivo* pode ser mantida simples ou cada vez mais pormenorizada até que finalmente se torne uma descrição totalmente estruturada. Este método de progressão é extremamente valioso no ensino. Também na coreografia quando apenas a essência do movimento precisa declarar.

Descrição Esforço-Forma é o termo aplicado para a investigação do movimento de acordo com o seu conteúdo dinâmico A palavra *Esforço* refere-se ao uso da energia. Este método de observação e análise, e os seus símbolos estão interessados na mudança de padrões que ocorrem no fluxo e refluxo de energia dentro do corpo. *Forma* refere-se especificamente à expressividade inerente na forma que o movimento toma. O observador deve observar a relação do caminho do gesto para o intérprete ou para a direcção dimensional, a fim de avaliar o seu valor expressivo ou funcional. Juntos, *Esforço-Forma* fornecem uma descrição valiosa do movimento em termos das suas qualidades e expressões, em contraste com a *Descrição Estrutural* que tem sido padronizada na Notação Laban. A aproximação *Esforço-Forma* é particularmente valiosa na área da fisioterapia e psicoterapia, em avaliação pessoal, e na

indústria. Tem também um grande contributo na antropologia e atletismo assim como em todos os tipos de dança.

Descrição Estrutural é o termo dado à descrição do movimento em termos mensuráveis e claramente definidos. Tal descrição, a mais usada geralmente, expressa movimento em termos de:

- *Corpo* - partes específicas que se movem.
- *Espaço* - direcção, nível, distância ou grau do movimento específico.
- *Tempo* - métrica e duração, como uma nota musical, uma semicolcheia.
- *Dinâmica* - a qualidade ou textura do movimento, isto é, se é forte, pesado, elástico, acentuado, enfatizado,etc.

A motivação para o movimento pode vir de várias fontes: direcção de destino, movimento, mudança anatómica, desenho visual, relação, centro de massa e equilíbrio, dinâmica e padrão rítmico. A *Descrição Estrutural* está mais preocupada com a direcção do destino, ou seja, para onde o corpo vai no espaço.

A Figura 2.3 mostra como a Análise de Movimento Laban pode ser definida assim em elementos básicos e irredutíveis, sob a forma de componentes.

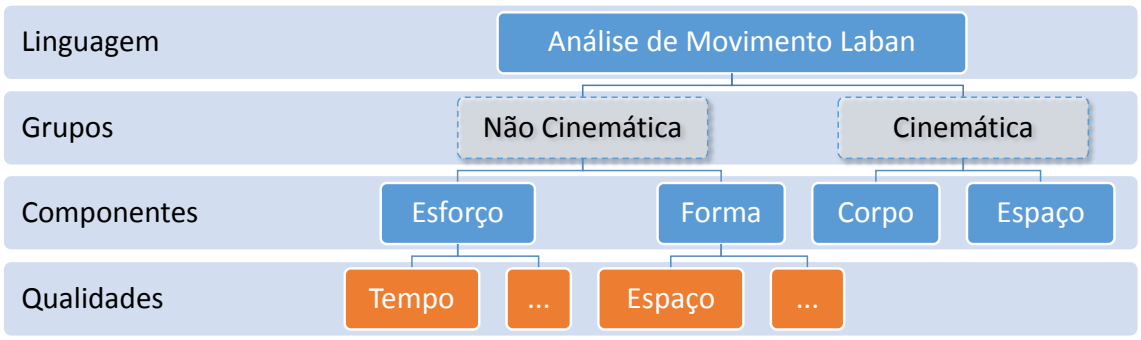


Figura 2.3: Fluxograma de Análise de Movimento Laban

2.2.1 Componente Corpo

A categoria *Corpo* descreve então as características estruturais e físicas do corpo humano e movimento. Esta categoria é responsável por descrever quais as partes do corpo que se estão a mover, que partes estão ligadas, que partes são influenciadas por outros e informações gerais sobre a organização do corpo. Algumas sub-categorias de *Corpo* são:

- Início do movimento a partir de órgãos específicos;
- Ligação dos corpos diferentes uns dos outros;

- Sequenciação de movimento entre as partes do corpo;
- Padrões de organização do corpo e conectividade

2.2.2 Componente Espaço

A categoria *Espaço* envolve movimento em conexão com o meio ambiente e com padrões espaciais. Esta categoria também descreve e anota as escolhas que se referem especificamente ao espaço, prestando atenção a:

- *Cinesfera* - a área que delimita o espaço onde o corpo se move e como o intérprete presta atenção a isso.
- *Intenção Espacial* - direcções ou pontos no espaço que o intérprete identifica ou usa.
- *Observações geométricas* - para onde o movimento está sendo feito, em termos de ênfase de direcções, lugares no espaço,

Os componentes *Corpo* e *Espaço* concentram-se assim em propriedades geométricas de movimentos de partes do corpo, em que os seus movimentos espaciais são movimentos relativos. Neste trabalho, estudaram-se as duas componentes que tratam a expressividade do movimento, *Esforço* e *Forma*.

2.2.3 Componente Esforço

A componente *Esforço* trata das propriedades dinâmicas do movimento no que toca a intenções internas. É dividida em quatro diferentes qualidades: *Espaço*, *Tempo*, *Peso* e *Fluxo*. Cada qualidade tem dois estados mutuamente exclusivos, como se pode ver na Tabela 2.1. Neste caso, irá descartar-se a qualidade *Peso* já que esta é habitualmente associada a força e, para o nosso caso não são adequadas para a sua caracterização.

Factor de Esforço	Elementos de Esforço	
Espaço	Direct	Indirect
Peso	Strong	Light
Tempo	Sudden	Sustained
Fluxo	Bound	Free

Tabela 2.1: Qualidades da Componente *Esforço*

Teoricamente, num movimento normal de uma pessoa, apenas são observadas três das quarto qualidades de Esforço, sendo muito díficil, mesmo para o mais experiente, executar acções combinando as quatro. Cada combinação de três categorias dá origem à qualificação de um gesto, por exemplo, uma gesto combinando as qualidades de *Espaço*, *Peso* e *Tempo* é considerado como acção de Esforço. Neste trabalho assume-se uma simplificação, ao considerar-se qualquer acção como tendo presentes as qualidades de *Espaço*, *Tempo* e

Fluxo, pois como atrás referido, a qualidade *Peso* é dificil de ser associada a características visuais.

2.2.4 Componente Forma

A componente *Forma* emerge das componentes *Corpo* e *Espaço*. Como o nome indica, trata a forma geométrica que o corpo e como esta muda ao longo do tempo É usada para integrar diferentes categorias em movimento. Esta também se divide em qualidades: *Forma*, Direcional e *Qualidade*

- *Forma Forma*, como o nome indica, é a forma estática que o corpo toma, que é maioritariamente geométrica É composto por três estados sendo eles *Wall-like*,*Ball-like* e *Pin-like*. Simplificando o espaço dos estados para o nosso caso, considera-se *Wall* e *Pin* como um único estado, onde o corpo está predominantemente levantado.

- *A Forma Direcional* representa a relação existente entre o corpo e o meio ambiente. Divide movimentos em *Spoke-like* (exemplo de apontar) e *Arc-like* (exemplo de acenar). É maioritariamente geométrico.

- *Qualidade de Forma* descreve a extensão do corpo ou como muda de forma em relação a orientações espaciais específicas. Termos mais específicos incluem *Rising* e *Sinking* (ao logo do eixo vertical do corpo), *Spreading* e *Enclosing* (ao longo do eixo horizontal), e *Advancing* e *Retreating* (ao longo do eixo sagital). Características geométricas são importantes, mas também formam alterações no tempo, em relação ao centro do corpo.

2.3 Espaço de Estados

A partir da descrição anterior, estamos em condições de desenvolver o espaço de estados para as componentes do modelo de Análise de Movimento Laban, e consequente implementação do nosso sistema de classificação.

$$\mathcal{L} = \begin{cases} c_1 \text{ :Esforço Tempo} & \in \{sudden, sustained\} \\ c_2 \text{ :Esforço Espaço} & \in \{direct, indirect\} \\ c_3 \text{ :Esforço Fluxo} & \in \{free, bound\} \\ c_4 \text{ :Forma Forma} & \in \{wall/pin, ball\} \\ c_5 \text{ :Forma Direcional} & \in \{spoke, arc\} \\ c_6 \text{ :Forma X} & \in \{spreading, enclosing\} \\ c_7 \text{ :Forma Y} & \in \{rising, sinking\} \\ c_8 \text{ :Forma Z} & \in \{advancing, retreating\} \end{cases} \tag{2.1}$$

Como apresentado na Figura (2.1), são enumeradas 8 qualidades pertencentes às componentes de Laban, Esforço e Forma. O espaço de estados, tem a sua origem nas definições da notação Laban, onde são definidos estados mutuamente exclusivos. A excepção é feita a c_4,

onde se assumem, por motivos de simplificação, que os estados *Wall* e *Pin* são usados para caracterizar uma pessoa cuja sua pose é estar de dominamente de pé.

Capítulo 3

Processamento e Extracção de Silhuetas

O processamento de imagem é uma área de investigação onde se estuda o desenvolvimento e aplicação de vários algoritmos com o objectivo de melhorar a qualidade das imagens, realçar características ou extrair vários tipos de informação. O processo passa geralmente por uma fase inicial de remoção de ruído, seguido da aplicação de algoritmos de segmentação, para que na imagem resultante sejam de fácil identificação os artefactos de interesse, como por exemplo objectos de determinada cor ou silhuetas de pessoas. Numa primeira instância, usa-se um qualquer sensor de visão de modo a obter-se uma imagem. Esta pode apresentar diversas imperfeições com origem no sensor ou na transmissão dos dados, como por exemplo a presença de pixeis ruidosos, contraste ou brilho inadequados. Além dos problemas originados aquando a aquisição, existem também outro tipo de desafios, como são exemplo as oclusões parciais ou totais dos artefactos de interesse, e outros problemas inerentes ao contexto dinâmico da cena a observar. Neste trabalho, assume-se, em geral, a presença de somente uma pessoa num cenário desobstruído. Desta forma, foi-nos permitido concentrar esforços na racionalização e representações alternativas das características visuais, como base do modelo Bayesiano para a Análise de Movimento de Laban.

3.1 Processamento da silhueta

Com o intuito de modelar a Análise de Movimento de Laban com base em características visuais, optou-se pela utilização de uma característica visual clássica, silhuetas. Para a sua obtenção, recorreu-se ao uso de algumas técnicas, cuja popularidade na área do processamento de imagem é sobejamente reconhecida. Como foi referido, as imagens podem conter vários focos de ruído, nomeadamente a qualidade da câmara ou problemas na transmissão de dados até à unidade de processamento. Dois tipos de ruído muito comuns são o *salt and pepper* e a existência de pixeis sem informação. De modo a mitigar o seu efeito, vamos aplicar um filtro de mediana. Este é um filtro de vizinhança, onde o valor de um pixel é substituído pela mediana dos pixeis da sua vizinhança.

Após remoção do ruído, vamos então concentrar-nos na subtracção de imagens. O objectivo principal da técnica da subtracção de imagens é realçar os pixeis cuja intensidade seja diferente em diferentes instantes de tempo. Se a perspectiva da câmara for estática, a subtracção de uma imagem por outra resulta na remoção de áreas cujas intensidades se-

jam semelhantes, deixando vincadas as zonas onde elas são diferentes, isto é, zonas onde se evidenciam as diferenças. A subtracção de imagens obedece à seguinte equação:

$$I_s = I_b - I, \tag{3.1}$$

onde I_b é uma imagem que representa todos os elementos estáticos da imagem, e I a imagem onde potencialmente se encontram objectos de interesse. Sendo (3.1) uma operação matricial, torna-se trivial que a subtracção de duas imagens iguais resulta numa matriz de zeros, ou seja, uma imagem preta e sem qualquer informação de interesse. Esta técnica apresenta resultados qualitativos quando usada em meios estáveis, ou seja, em locais onde o ambiente natural capturado se mantém praticamente igual. Alterações bruscas das condições de luminosidade condicionariam claramente este método. Existem algumas técnicas adaptativas, que permitem ir actualizando a informação das características estáticas, de modo a mitigar estes efeitos. No nosso caso, tendo em conta que o cenário é idealizado para ambientes controlados, (câmara fixa no interior do edifício), assume-se que as condições de aquisição se mantêm estáveis ao longo do tempo. Com isto tudo torna-se evidente que a presença de uma pessoa pode ser facilmente identificada em termos computacionais através da subtracção das características estáticas (Figura 3.1).

Figura 3.1: Resultado (I_s) da subtração de uma imagem (I) a uma imagem padrão (I_b) contendo o seu fundo estático.

Após obtermos uma imagem, cujo conteúdo se espera reter maioritariamente objectos de interesse (como se verifica na imagem mais à esquerda da Figura 3.1, vamos aplicar um algoritmo que irá permitir a extracção do contorno de silhueta de uma pessoa. Para o efeito pode aplicar-se o método de Canny. É provavelmente o mais famoso operador de detecção de contornos em imagens, e foi concebido com 3 objectivos principais:

1. Detecção óptima.
2. Boa localização, minimizando o erro entre contorno calculado e real.
3. Evitar respostas ambíguas.

A primeira etapa consiste na remoção de ruído. Para tal, Canny foi o primeiro a demonstrar que o operador Gaussiano era óptimo na detecção de arestas. A segunda etapa foca-se na precisão, o que é conseguido à custa de um processo de supressão não máxima. Este processo

retém os pontos máximos de um aglomerado de pontos. O último objectivo, resulta na localização de um ponto pertencente a uma aresta em reposta a uma mudança de intensidade luminosa. Encontrando o gradiente de intensidade de uma imagem , é procurado o máximo local na direcção do gradiente de modo a se encontrarem as arestas. O resultado final é uma imagem binarizada, onde os pixeis a branco definem os contornos encontrados através do método de Canny.

No nosso caso particular, partindo do princípio que a remoção das zonas estáticas é bem sucedido, a detecção de contornos fechados com o algoritmo de Canny, facilmente irá retornar um contorno fechado da silhueta da pessoa a observar. Na Figura 3.2 ilustra-se de forma

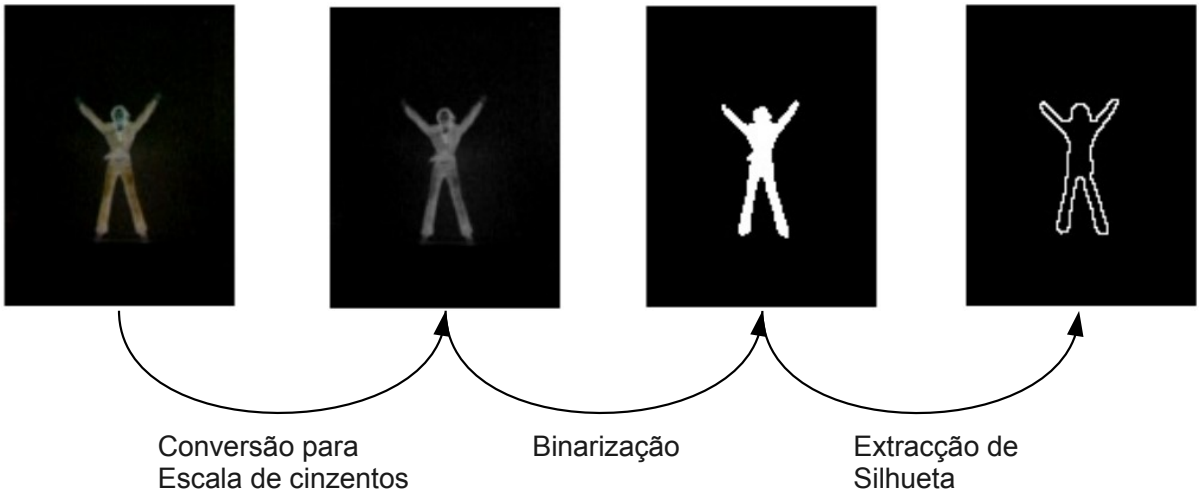


Figura 3.2: Ilustração simplificada do processo de extracção de silhuetas a partir de uma imagem de subtração.

simplificada as etapas de processamento básicas para se obter uma silhueta. O algoritmo de Canny corre geralmente sobre imagens em escala de cinzentos. Contudo, conforme se pode verificar, a simples binarização e erosão de resultante binária pode permitir, de forma simples, a extracção de uma silhueta. Temos de ter em atenção que estes métodos só são aplicados com sucesso em situações em que as condições de aquisição sejam controladas.

3.2 Variáveis de Representação de Características

Considere-se a imagem binária I_{BW} obtida na secção anterior, contendo apenas o contorno de uma silhueta P, representadas por:

$$P = \begin{bmatrix} (u_1, v_1) \\ \vdots \\ (u_s, v_s) \end{bmatrix}, \forall\ (u_s, v_s) \in I_{BW} : I_{BW}(u_s, v_s) = 1. \tag{3.2}$$

Dada a potencial alta dimensão de P, vamos usar uma representação alternativa baseada no procedimento algorítmico da Análise de Componentes Principais (PCA). Este é um método que tem como finalidade original, a análise de dados com o propósito de os reduzir, eliminando

sobreposições e escolhendo uma forma mais representativa dos dados a partir de combinações lineares das variáveis originais. Contudo, neste trabalho irá usar-se apenas a informação que define o espaço das componentes principais.

3.2.1 Componente Estática - Análise das Componentes Principais

Análise de Componentes Principais (PCA) é um método que permite identificar padrões num conjuntos de dados, e expressar todos esses dados de maneira a destacar as suas semelhanças e diferenças. Em dados multi-dimensionais onde a representação gráfica não é possível, padrões podem ser difíceis de ser identificados. Nestes casos PCA torna-se numa poderosa ferramenta para a sua análise. A grande vantagem de PCA é que além da identificação de padrões, permite a compressão dos dados reduzindo a sua dimensão, minimizando a perda de informação. Trata-se de uma representação estatística única de um conjunto de pontos, onde a componente principal identifica a orientação que maximiza a variância dos dados. A informação das componentes é aqui colocada como hipótese de representação de características geométricas. Usando os valores das componentes principais (valores e vectores próprios da matriz de covariância dos dados), torna-se possível representar alternativamente um silhueta através da caracterização de padrões, e simultaneamente obter a desejada redução dimensional dos dados. Para uma melhor compreensão explica-se a técnica faseadamente.

- Passo 1: Obter um conjunto de dados, neste caso, uma imagem binária, equivalente a um gráfico de coordenadas x,y em que as coordenadas dos pontos são as coordenadas dos pixeis de valor igual a 1.

- Passo 2: Subtrair a média ao conjunto de pontos, i.e., calcular a média em cada dimensão e subtrair a todas as coordenadas dessa dimensão esse valor. O resultado será uma distribuição de média zero.

- Passo 3: Calcular a matriz de covariância da amostra resultante.

- Passo 4: Calcular os vectores e valores próprios dessa matriz. A maioria do software disponível para o cálculo de PCA já inclui essas rotinas implementadas.

- Passo 5: Armazenar as coordenadas de vectores e valores próprios. O maior valor próprio e o vector associado definem a componente principal, o outro define a componente secundária. São estes valores que definem a distribuição no espaço de PCA.

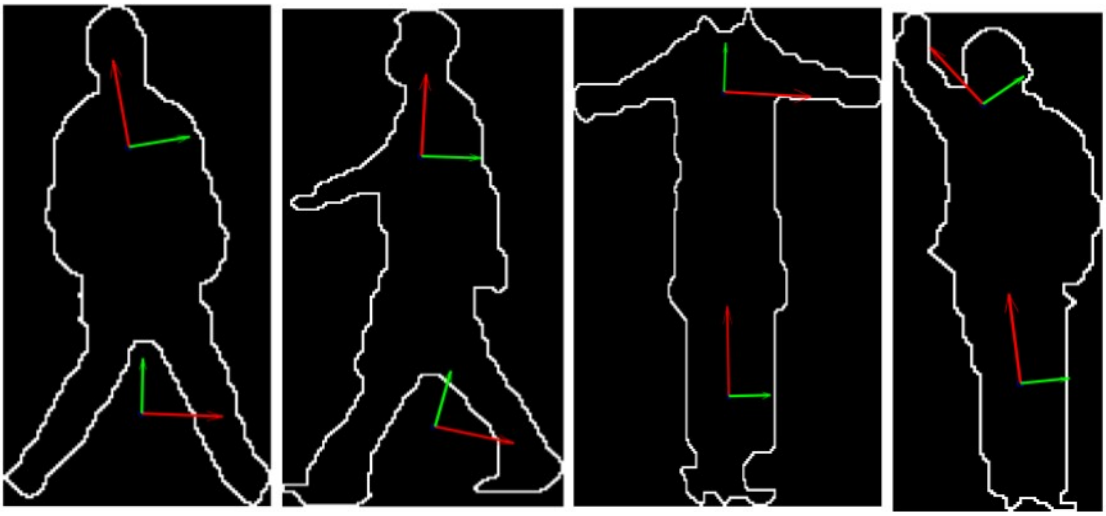
No fundo, com a aplicação desta técnica, pretende-se determinar os parâmetros da elipse que melhor se ajusta à nuvem de pontos que define a silhueta, onde o eixo maior da elipse é dado pela componente principal.

Na prática, são usadas as coordenadas dos vectores próprios $v_r = (x_r, y_r)$ como características de silhuetas, para cada imagem estática I^n_{BW}, numa dada amostra n. Calculou-se o rácio entre o primeiro e segundo valor dos valores próprios, λ_1/λ_2, multiplicando-se depois

este valor pelo primeiro termo dos vector próprio, de tal forma que $v_1' = v_1(\lambda_1/\lambda_2)$. O segundo vector próprio usa-se na sua forma normalizada. Esta etapa de processamento coloca o primeiro vector próprio com informação implícita dos valores próprios, mitigando assim o impacto da escala nas silhuetas obtidas.

$$\hat{P}^n = pca(P) : \hat{P} = [v_1' \; v_2] \equiv [p_1^n, \cdots, p_4^n]. \tag{3.3}$$

A técnica do PCA é aplicada por duas vezes, onde a silhueta da imagem é dividida em dois, separando-se em partes superior e inferior do corpo através da informação do centro de massa. Justifica-se este passo pela seguinte razão. Na maioria dos movimentos naturais, executados por uma pessoa que esteja de pé, o rácio entre altura e largura de uma silhueta vai ser muito semelhante, qualquer que seja o movimento, de tal forma que a aplicação da



(a) Exemplo de características PCA para diferentes gestos.

(b) Exemplo de características PCA para o mesmo gesto, sobrepostos na imagem original.

Figura 3.3: Exemplos de caractísticas PCA no espaço de silhueta, e sobrepostas nas imagens originais.

técnica de PCA à silhueta como um todo, apresentou dados pouco discriminantes. A solução encontrada foi a de dividir a silhueta em duas partes, cada uma caracterizando um hemisfério diferente do corpo humano. Desta forma em situações em que o movimento é dominado por uma das partes do corpo, a aplicação de PCA vai revelar alterações significativas, quer a nível de valores como de vectores próprios, obtendo-se assim uma melhor caracterização geométrica. A estratégia é devidamente ilustrada na Figura 3.3.

3.2.2 Componente Dinâmica - Transformada de Fourier

Contudo, algumas das componentes da LMA caracterizam a expressividade do movimento, intimamente ligadas a propriedades dinâmicas, propriedades estas que são melhor caracterizadas se analisarmos sequências de imagens, ao invés de apenas uma. Olhando para as componentes de *Esforço*, estas necessitam de uma avaliação ao longo do tempo, o que não é possível ser feito através da técnica PCA.

Assim considere uma sequência de imagens $\mathbf{I}$ dividida em sub-sequências com um número de amostras $\hat{n}$, onde $\hat{\mathbf{I}} = \{I_{n-\hat{n}}, \cdots, I_n\}$. Para cada $\hat{I}$ tem-se uma série temporal discreta para cada característica p_j, sendo que $p_j[\hat{n}] = (p_j^{n-\hat{n}}, \cdots, p_j^n)$. Com isto, $p_j[\hat{n}]$ constrói um sinal temporal de tamanho $\hat{n}$. Para caracterizar a **dinâmica do movimento** decidiu aplicar-se a Transformada de Fourier. A Transformada de Fourier é um método bastante usado para estudar e caracterizar o comportamento de sinais ao longo do tempo. No domínio discreto é aplicado método da Transformada Discreta de Fourier (DFT). Esta converte um número finito de amostras de um sinal em coeficientes de uma combinação finita de sinusóides, ordenados pela sua frequência.

No entanto o cálculo da DFT costuma ser realizado por um algoritmo mais eficiente, chamado Transformada Rápida de Fourier. A Transformada Rápida de Fourier (FFT) é um método matemático muito eficiente que reordena os cálculos dos coeficientes da (DFT). Trata-se de um algoritmo que realiza uma avaliação da DFT com um menor esforço computacional, ao invés de realizar o cálculo da DFT directamente pela sua definição.

Representa-se assim $p_j[\hat{n}]$ no domínio da frequência através de $P_j(\omega)$, onde os coeficientes calculados representam um conjunto de características, contendo implicitamente informação sobre a dinâmica de $\hat{\mathbf{I}}$.

$$P_j(\omega) = \sum_n p_j[\hat{n}] e^{-i\omega n}. \tag{3.4}$$

Mais especificamente, para o conjunto dos coeficientes de Fourier, seleccionou-se o valor máximo, tal que

$$F = \{f_1, \cdots, f_j\} \ : \ f_j \in \max P_j(\omega), \tag{3.5}$$

e o seu índice da frequência fundamental correspondente.

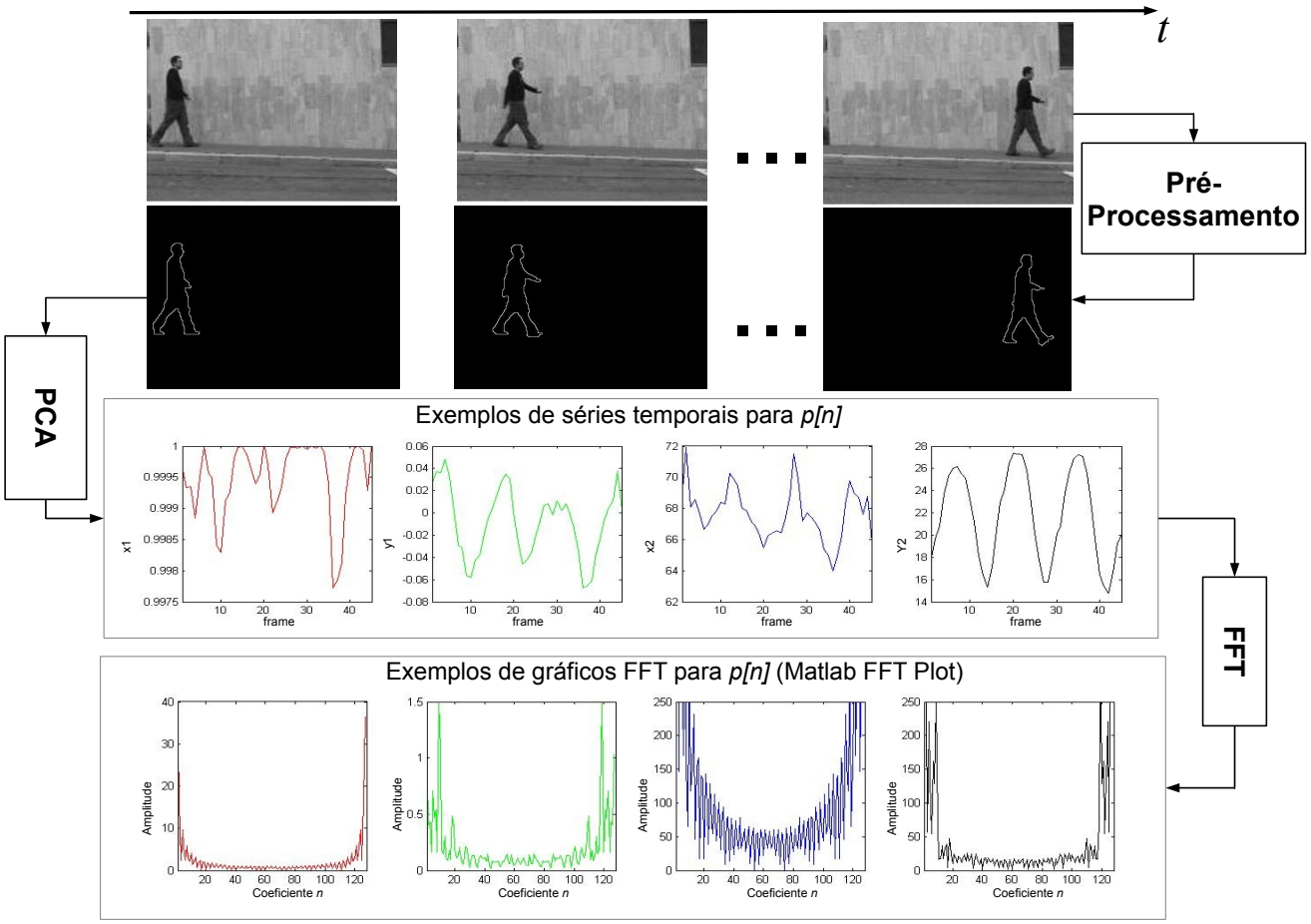


Figura 3.4: Diagrama da sequências de processos $\hat{\mathbf{I}}$.

3.2.3 Característica Complementar - Orientação

A última característica considerada refere-se ao vector deslocamento da silhueta $\overrightarrow{d}$, para duas imagens consecutivas, tal que

$$\overrightarrow{d}_n = (u_c^n - u_c^{n-1}, v_c^n - v_c^{n-1}), \tag{3.6}$$

onde (u_c^n, v_c^n) corresponde ao centro de massa para a amostra n. A característica deslocamento é então dada por

$$\theta_n = atan2(v_c^n - v_c^{n-1}, u_c^n - u_c^{n-1}), \tag{3.7}$$

que será especialmente relevante para componentes de LMA que se baseiam na direcção, em vez da orientação. Contudo, esta variável, é apresentada como sendo uma informação complementar à que resulta das características geométricas de PCA.

A Figura 3.4 ilustra a sequência de processos realizados neste capítulo, desde a obtenção da silhueta até à extracção das características estáticas e dinâmicas. Considere-se uma sequência temporal de imagens, onde é possível observar uma sequência de acções. A cada nova imagem, esta é processada de modo a que se extraia a silhueta da pessoa em movimento.

As características estáticas (com base em PCA) são extraídas para cada imagem. Na terceira linha de imagens, podemos observar uma sequência temporal de características PCA ao longo da sequência apresentada. A cada instante, o valor da série temporal representa uma característica estática. Na última linha, apresentam-se características dinâmicas, isto é, os gráficos ilustrativos da série de Fourier para os gráficos correspondentes da linha anterior.

3.3 Resultados Experimentais

Nesta secção complementam-se os resultados experimentais apresentados anteriormente para cada uma das diferentes componentes. Calculou-se um conjunto de amostras para um conjunto representativo de características, mais especificamente $\{x'_1, y'_1, {}^{\lambda_1}/_{\lambda_2}, \} \in \hat{P}$,θ e $\{f_1, f_2\} \in F$, para diferentes gestos g_{index} nos datasets seleccionados. A Figura 3.5 apresenta a média dos valores medidos para cada característica em cada um dos diferentes gestos. Os resultados mostraram que as características são discriminantes em relação aos diferentes

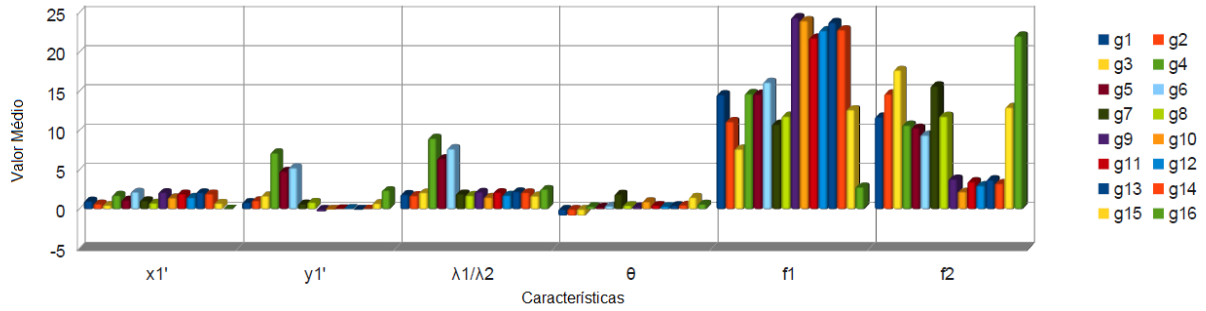
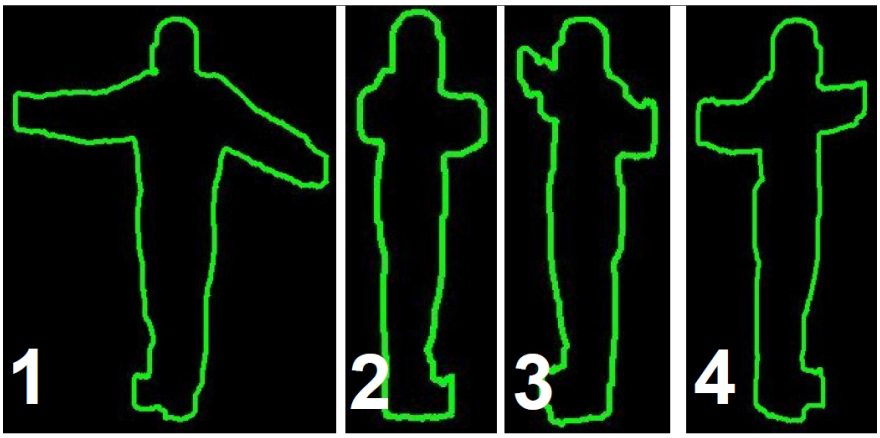


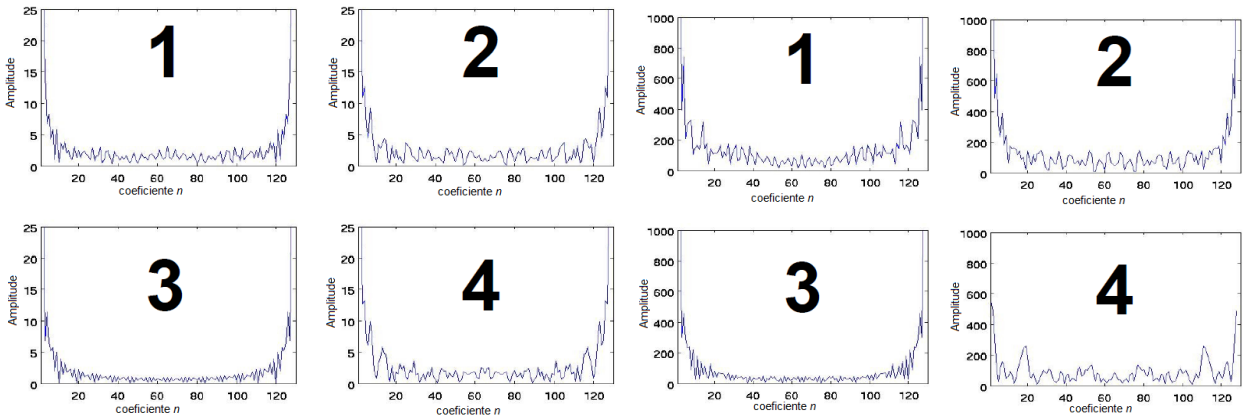
Figura 3.5: Valores médios para as características $\{v'_1, {}^{\lambda_1}/_{\lambda_2}, \theta, f_1, f_2\}$, calculados para 16 gestos g, relativos aos datasets usados neste trabalho.

gestos, e consequentemente, diferentes propriedades do movimento executado. As diferenças ilustradas indicam que as características seleccionadas contêm efeitos discriminatórios, representativos de propriedades geometrias e dinâmicas, a serem exploradas pela formulação do nosso modelo Bayesiano de LMA, no Capítulo 2. Apresentamos também um propriedade interessante observada durante um estudo preliminar na observação de acções de perspectivas diferentes.

A última imagem da Figura 3.6a mostra que as características geométricas são naturalmente diferentes, contudo o comportamento da sua FFT apresenta-se semelhante nas várias perspectivas 3.6b e 3.6c. Isto é um indicador a ser explorado no futuro, por forma a robustecer a geração de características discriminantes, a incorporar numa versão estendida do modelo.



(a) Silhueta de um sujeito para 4 perspectivas diferentes, no instante n.



(b) Característica p_1

(c) Característica p_3

Figura 3.6: Diferentes FFT para as características p_1 e p_3, em 4 perspectivas diferentes.

Capítulo 4

Modelo Bayesiano para Análise de Movimento Laban

Reconhecimento de padrões é uma disciplina científica cujo objectivo é a classificação de informação em categorias ou classes. Dependendo da aplicação, essa informação pode ter origem em vários tipos de sinais sobre os quais se queira proceder à segmentação (ou separação em categorias) de informação através de um processo de classificação. Apesar de ser uma área com grande tradição na investigação, só depois da segunda metade do século XX é que se popularizou como disciplina de investigação. Isto porque na passagem da fase industrial para a pós-industrial, a necessidade de tratamento e recuperação de informação tornaram-se cada vez mais importantes na automatização de máquinas e processos. O reconhecimento de padrões é, hoje em dia, uma área científica de grande importância no que diz respeito à investigação e a aplicações na engenharia. Assim sendo, esta disciplina de investigação encontra-se presente na maioria dos sistemas computacionais inteligentes desenhados para realizar decisões autonomamente, sendo que a visão por computador é uma área bastante abrangente desse tópico. Um sistema captura imagens através de uma câmara, para que estas sejam posteriormente analisadas, gerando assim diversos tipos de informação, como por exemplo detecção de objectos específicos. Um exemplo simples é a inspecção visual numa linha de montagem de uma fábrica industrial. As imagens são geralmente analisadas em tempo real e o sistema de reconhecimento classifica se existe ou não algum defeito no fabrico, num processo em que geralmente se comparam os dados observados com um modelo pré existente, num processo de classificação. A métrica de comparação entre o que é observado e o modelo, revela a confiança que o sistema tem na estimação do estado de um determinado objecto. Existem diversos tipos de classificadores, exemplos de *Redes Bayesianas*, *K-Nearest Neighboors* ou *Support Vector Machines*, sendo que todos eles demonstram na literatura, a capacidade de classificar com elevadas taxas de sucesso.

Neste trabalho usou-se um classificador Bayesiano para avaliar o modelo desenvolvido no que diz respeito às suas capacidades de análise de movimento. Para que a análise seja considerada tempo real, o modelo tem que ser capaz de caracterizar com precisão sequências de pequena duração.

Os modelos Bayesianos são frequentemente representados por modelos gráficos, onde os nós representam variáveis, e as setas representam dependências condicionais. Modelos com-

pletos de variáveis e suas relações podem ser usadas para responder a questões probabilísticas sobre elas, ou seja, estimar o estado de uma variável, condicionada ao conhecimento de novas observações, estimativas prévias e de um modelo (geralmente construído através de dados experimentais previamente disponíveis). Para a especificação completa da rede Bayesiana, é necessário representar para cada nó, a distribuição de probabilidade condicionada do nó para as suas dependências, $P(nó\ /dependência)$. Existem várias métodos probabilísticos que são baseados no teorema de Bayes, e que têm a uma representação equivalente sob a forma de Redes Dinâmicas Bayesianas [Murphy, 2002].

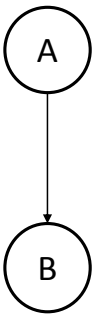


Figura 4.1: Modelo gráfico

O teorema de Bayes estabelece o grau de confiança sobre um estado de uma variável A após novas evidências B serem conhecidas, e cujo formalismo mais simples se apresenta conforme a equação seguinte.

$$P(A|B) = P(A) \times \frac{P(B|A)}{P(B)} \tag{4.1}$$

O termo $P(A|B)$ define a probabilidade à posteriori. O conhecimento que se tem sobre A antes de novas evidências serem tidas em conta, é dado pela probabilidade à prior $P(A)$. A existência deste tipo de distribuição é o que estabelece as diferenças entre aproximações frequentistas e subjectivistas (ou Bayesianas). Além do seu significado, matematicamente ela tem o objectivo de regularizar a distribuição à posteriori, ou seja, que se previna o *sobreajuste*, se proporcione a existência de uma solução e que esta seja única.

O termo $P(B|A)$ define a probabilidade da evidência, que representa o modelo probabilístico do fenómeno a analisar. É normalmente definido a partir de dados experimentais, num processo em que se pretende estimar uma densidade de probabilidade em função da variável evidência B, dado o conhecimento do estado de A.

As perguntas ao modelo (distribuição a posteriori) são respondidas através de Inferência Bayesiana. Existem vários métodos de inferência. Para este caso usamos aquele que é talvez o método mais popular, denominado de *Maximum a Posteriori (MAP)* que tem por base a formulação de Bayes. Inferência usando MAP, é um método de estimação pontual que determina o valor da variável que se pretende estimar, maximizando a densidade da distribuição à posteriori, usando a regra de Bayes, tal que:

$$A_{MAP}(B) = arg_A max P(B|A)P(A), \tag{4.2}$$

onde B representa a evidência, e A a variável a ser estimada, definidas anteriormente.

4.1 Programação Bayesiana

Para o desenvolvimento de modelos Bayesianos existe um formalismo intuitivo que permite uma implementação metódica e eficiente, denominado Programação Bayesiana [Bessiere et al., 2012]. Este formalismo foi desenvolvido por forma a juntar 2 mundos: o matemático e o da programação. Dentro deste espírito, a solução de um modelo Bayesiano fica pronta à sua rápida implementação num programa de computador. Destaca-se ainda que este formalismo pode ser aplicado à generalidade dos algoritmos Bayesianos, facilitando assim a sua comparação. A programação Bayesiana é estruturada em 4 passos essenciais, os quais se enumeram de seguida, acompanhados com um pequeno exemplo didático, com base no modelo que se propõe nesta tese.

1 **Definição de Variáveis**: Escolha de variáveis relevantes para a implementação do modelo, respeitantes à informação disponível bem como ao que se deseja estimar.

EXEMPLO $\longrightarrow$ passo 1
Neste pequeno exemplo, pretende-se estimar o estado da componente *Esforço Tempo* com base na informação dada pela primeira característica de Fourier. Tem-se então 2 variáveis: • $c_1 \equiv \{sudden, sustained\}$: variável aleatória que representa a componente *Esforço Tempo* que pode tomar 2 estados possíveis. • $f_1 \in \mathcal{R}$: variável aleatória, cujo domínio é real.

2 **Decomposição**: Decompõe a distribuição conjunta de todas as variáveis identificadas, num produtório de probabilidades condicionadas mais simples, respeitando as dependências definidas.

EXEMPLO $\longrightarrow$ passo 2
O nosso modelo exemplo tem duas variáveis que definem a seguinte distribuição conjunta dada por $P(c_1, f_1)$. Se nos auxiliarmos na Figura 4.1, e considerarmos as variáveis c_1 e f_1 como A e B respectivamente, obtemos a seguinda decomposição, $$P(c_1, f_1) \propto P(f_1|c_1)P(c_1),$$ onde $P(c_1)$ e $P(f_1|c_1)$ definem as distribuições a priori e de evidência respectivamente. Apesar de neste caso, o modelo gráfico ser simples, a sua utilidade revela-se à medida que o número de variáveis aumenta, permitindo uma estruturação e um fácil seguimento das dependências entre nós.

3 **Formulação**: Nesta etapa, decide-se que tipo de distribuição é adequada para representar a distribuição da decomposição.

EXEMPLO $\longrightarrow$ passo 3

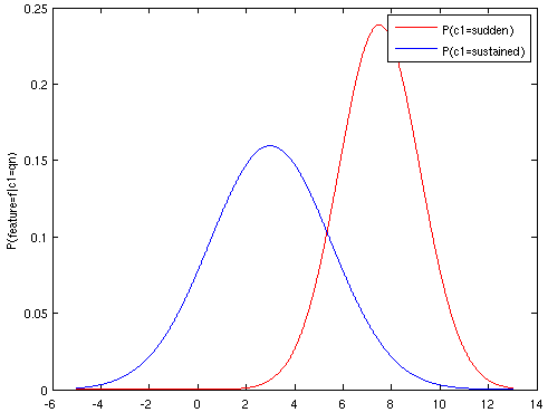
A decomposição proposta no passo anterior revela 2 distribuições. Neste exemplo, vamos assumir que estamos a observar um movimento, que sabemos ser tendencialmente *sudden*. Então, vamos com base nesta informação definir uma distribuição à priori $P(c_1)$ que simplesmente considera um histograma, cuja probabilidade à priori é dada por,

$$P(c_1) = \begin{cases} P(c_1 = sudden) = 0.6 \\ P(c_1 = sustained) = 0.4 \end{cases}$$

A distribuição de evidência $P(f_1|c_1)$ é uma distribuição probabilística, onde a densidade é estimada em função da variável aleatória f_1. Para o nosso exemplo, consideramos simplesmente $f_1 \in \mathcal{R}$, pelo que é comum definir-se uma densidade do tipo Gaussiana. Assuma que existem os seguintes vectores de amostras para f_1.

$$\text{amostras } f_1 = \begin{cases} [6, 7, 8, 9], \text{ for } c_1 = sudden \\ [1, 2, 3, 4, 5], \text{ for } c_1 = sustained \end{cases}$$

Destas amostras é trivial calcular as médias e variância para cada um dos estados possíveis. Desta forma, obtém-se a distribuição de evidência seguinte:



Onde os parâmetros calculados resultam nas distribuições seguintes:

$$P(f_1|c_1) = \begin{cases} \mathcal{N}(7.50, 1, 67), \text{ for } c_1 = sudden \\ \mathcal{N}(3.00, 2.50), \text{ for } c_1 = sustained \end{cases}$$

Ficando desta forma, definida e concretizada a formulação das distribuições da nossa decomposição.

4 **Questão**: A questão formula uma pergunta ao programa, sobre o que se pretende estimar baseado no que se tem disponível para observar. A resposta à questão é dada, usando Inferência Bayesiana.

EXEMPLO $\longrightarrow$ passo 4

Uma vez definido o nosso modelo exemplo, estamos em condições de formular a questão. Queremos saber qual o estado mais provável para a componente *Esforço Tempo*, dada a existência de uma nova evidência, calculada a partir de uma sequência de imagens. Esta questão é formulada como $P(c_1|f_1)$. A nossa formulação permite ao nosso problema ser de forma fechada, de tal modo que a solução para a inferência é dada analiticamente. Considere, por motivos de simplificação matemática, que $sudden = q_1$ e $sustained = q_2$. Então, para um valor da evidência observada $f_1 = 4.5$ temos que:

$$\begin{aligned} P(c_1 = q_1|f_1 = 4.5) &= \frac{P(c_1{=}q_1)P(f_1{=}4.5|c_1{=}q_1)}{P(c_1{=}q_1)P(f_1{=}4.5|c_1{=}q_1)+P(c_1{=}q_2)P(f_1{=}4.5|c_1{=}q_2)} \\ &= \frac{0.6\times0.0476}{0.6\times0.0476+0.4\times0.1333} \\ &= 0.3488 \end{aligned}$$

e para o outro estado de c_1,

$$\begin{aligned} P(c_1 = q_2|f_1 = 4.5) &= \frac{P(c_1{=}q_2)P(f_1{=}4.5|c_1{=}q_2)}{P(c_1{=}q_1)P(f_1{=}4.5|c_1{=}q_1)+P(c_1{=}q_2)P(f_1{=}4.5|c_1{=}q_2)} \\ &= \frac{0.4\times0.1333}{0.6\times0.0476+0.4\times0.1333} \\ &= 0.6512 \end{aligned}$$

Então a distribuição à posteriori resulta em:

$$P(c_1|f_1 = 4.5) = \begin{cases} 0.3488, \text{ for } c_1 = sudden \\ 0.6512, \text{ for } c_1 = sustained \end{cases}$$

que, após aplicação da técnica de inferência MAP, resulta que o estado mais provável para o nosso exemplo é $c_1 = sustained$ com uma probabilidade de 65.12%.

Termina-se esta secção de introdução à programação Bayesiana, com o formalismo, tal como é definido em [Bessiere et al., 2012], do nosso programa para o exemplo apresentado, ilustrado na Figura 4.2.

Programa
- Descrição
  - Especificação
    - **Variáveis relevantes:**
      $c_1 \in \mathcal{L} : c_1 = q_1, q_2$: representa a componente Laban de *Esforço Tempo*;
      $f_1 \in F$: a 1^a característica dinâmica da silhueta;
    - **Decomposição:**
      $P(c_1, f_1) = P(c_1)P(f_1|c_1)$
    - **Formulação:**
      $P(c_n) = [P(c_1 = q_1) = 0.6, P(c_1 = q_2) = 0.4] \rightarrow$ **[Histograma]**
      $P(f_1|c_1) = \begin{cases} \mathcal{N}(7.50, 1, 67), \text{ para } c_1 = q_1 \\ \mathcal{N}(3.00, 2.50), \text{ para } c_1 = q_2 \end{cases} \rightarrow$ **[Gaussiana]**
  - **Identificação:**
    nenhuma;
- **Questão:**
  $P(c_1|f_1)$

Figura 4.2: Programa Bayesiano para o exemplo didático proposto na Secção 4.1.

4.2 Modelo para Componentes de Laban

As variáveis representando as características apresentadas em capítulos anteriores, são usadas para definir o modelo Laban, que sendo estas reais e contínuas (excepto quando mencionado) é modelado sob distrubuições Gaussianas. As diferentes componentes têm as suas dependências condicionadas a tipos de características adequadas, escolhidas heuristicamente com base em definições teóricas. O modelo de Análise de Movimento Laban proposto é uma representação probabilística das suas componentes, aprendida através de um conjunto de características visuais $\{f_i,\theta,p_i\}$, extraídas a partir de uma combinação de imagens estáticas e sequenciais. É parametrizado em vários sub-modelos, tantos quanto o número de componentes c_n, onde a hipótese inicial para os estados de espaço é definida na Figura (2.1), baseada nos conceitos da Secção 2.2. Das conclusões retiradas dos resultados da Secção 3.3, o modelo deve estar preparado para suportar um número arbitrário de câmaras S_m, tendo em mente que o modelo proposto é assumido correr de forma independente para cada sensor diferente. Na prática, cada câmara irá mostrar uma perspectiva diferente da silhueta, representada pela sua orientação relativa à posição da câmara.

4.2.1 Relações Componentes-Características

De seguida vamos, com base na Secção 4.2, associar heuristicamente os diferentes tipos de características a cada uma das componentes Laban analisas neste trabalho, nomeadamente a componentes *Esforço* e *Forma*.

c_1 *Esforço Tempo* caracteriza o processo cognitivo de decisão, que é justamente relacionado com o tempo. Portanto é associado a características dinâmicas f_i e a questão Bayesiana é dada por $P(c_1|f_i)$.

c_2 *Esforço Espaço* foca-se na atenção com respeito à *orientação com um propósito*. Coloca-se neste trabalho, a hipótese de ser uma combinação de características geométricas, dinâmicas e de orientação, tal que $P(c_2|f_i, p_j, \theta)$.

c_3 *Esforço Fluxo* caracteriza a continuidade do movimento, que é relacionado com a performance ao longo do tempo e se está ou não contido numa única acção. Dinâmica e deslocamento são os tipos de características seleccionadas, por isso $P(c_3|f_i, \theta)$.

c_4 *Forma Forma* é uma das categorias de componente *Forma*, e como o nome indica, é a forma que o corpo toma, que é maioritariamente geométrica, assim que $P(c_4|p_j)$. Simplificamos o espaço dos estados considerando *Wall* e *Pin* como um estado único, onde o corpo está predominantemente levantado.

c_5 *Forma Direcional* representa a relação existente entre o corpo e o meio ambiente. Divide movimentos em *Spoke-like* e *Arc-like*. É maioritariamente geométrico, portanto $P(c_5|p_j)$.

$c_{6,7,8}$ *Forma Espacial* também tem qualidades, que descrevem a extensão do corpo ou como muda de forma em relação a orientações espaciais específicas. Características geométricas são importantes, mas também formam alterações no tempo, em relação ao centro do corpo. Por isso, a questão Bayesiana é formulada como $P(c_6, c_7, c_8|f_i, p_j, \theta)$.

4.2.2 Programa Bayesiano

1. Identificação e definição das variáveis relevantes

As variáveis são definidas e baseadas na *Análise do Movimento Laban* que descreve de forma elementar características do movimento humano.

- $\mathcal{L} = \{c_n \equiv \{q_1, q_2\}\}$ é uma variável denotando uma componente Laban que representa a característica de um movimento específico, admitindo dois estados exclusivamente mútuos, como definido na secção 2.2. É assumido que um estado anterior observado se propaga dentro de uma sequência de movimento.

- $F = \{f_i \in \mathbb{R}_0^+\}$ são um conjunto de variáveis aleatórias que representam a informação dinâmica de uma sequência de imagens $\hat{I}$.

- $\hat{P} = \{p_j \equiv \{u_1, \cdots, u_x\}\}$ são um conjunto de variáveis aleatórias que representa informação geométrica de silhuetas estáticas $\in I$. É discretizado num número de intervalos equidistantes x, representando um possível estado u_x.

- $\theta \in [-\pi, \pi]$ é uma variável aleatória que representa a orientação do deslocamento entre duas silhuetas consecutivas, P^t e P^{t-1}.

2. Decomposição

A Decomposição tem como distribuição conjunta $P(Laban, Caracteristicas)$. De acordo com a regra de Bayes, a decomposição é iniciada pela camada inferior. No nosso caso particular, a decomposição é a seguinte:

- distribuição à priori $P(Laban)$;
- distribuição de evidência $P(Caracteristicas|Laban)$;
- distribuição à posteriori $P(Laban|Caracteristicas) \propto P(Laban) \times P(Caracteristicas|Laban)$, onde o sinal $\propto$ permite a omissão do factor de normalização.

A Figura 4.3 mostra o modelo gráfico proposto. Mais especificamente, temos como dis-

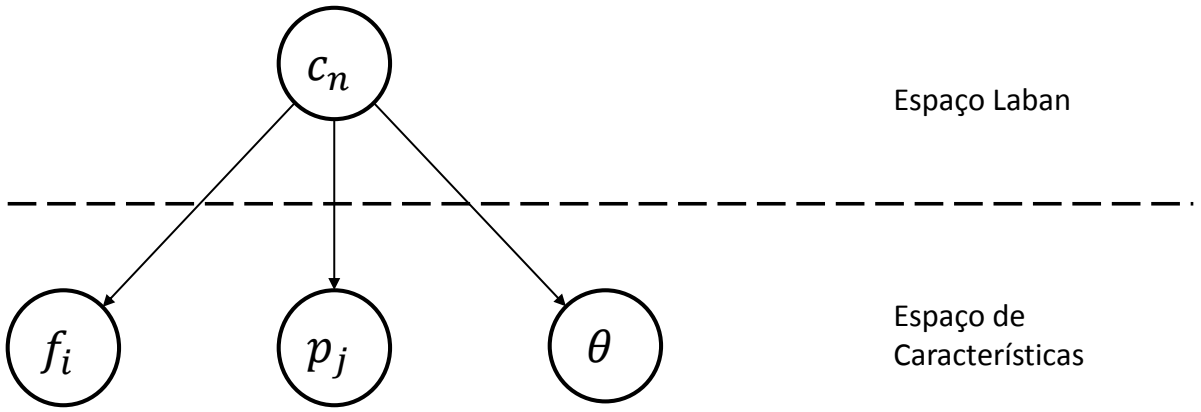


Figura 4.3: Modelo gráfico

tribuição conjunta $P(c_n, p_j, f_i, \theta)$, com a distribuição à priori $P(c_n)$ e a probabilidade de evidência $P(p_j, f_i, \theta|c_n)$. O termo $P(c_n|p_j, f_i, \theta) \propto P(c_n)P(p_j, f_i, \theta|c_n)$ define a distribuição à posteriori. O sinal $\propto$ aplica-se, pois por motivos de simplificação, geralmente omite-se o numerador (ou factor de normalização). Assim sendo, e aplicando recursivamente a regra da conjunção obtém-se a formulação completa para a decomposição, dada pela equação (4.3).

$$P(c_n, p_j, f_i, \theta) = P(c_n)P(p_1|c_n) \cdots P(p_j|c_n)P(f_1|c_n) \cdots P(f_i|c_n)P(\theta|c_n) \qquad (4.3)$$

3.Formulação

As variáveis de características são definidas como anteriormente descrito na secção 3.2. As probabilidades de evidência representam-se através de um modelo probabilístico que estabelece a relação entre as componentes Laban e a características presentes no espaço de características.
Os tipos de distribuições paramétricas do modelo proposto são dadas por:

- Distribuição à priori: $P(c_n) = uniforme; \qquad t = 0;$

- Distribuição à priori: $P(c_n) = P(c_n|p_j, f_i, \theta)$ calculada no instante anterior; $t \neq 0$;

- Distribuição de evidência: $P(p_j, f_i, \theta|c_n)$, Matriz estocástica;

- Distribuição à posteriori definida por:
  $P(c_n|p_j, f_i, \theta) \propto P(c_n) \prod_{\forall i,j} P(f_i, p_j, \theta|[c_n = q_j])$
  $\propto P(c_n)P(\theta|c_n) \prod_i P(f_i|c_n) \prod_j P(p_j|c_n).$

O facto de se definirem parametrizações com base em distribuições Gaussiana, permite que a inferência seja resolvida através de uma solução analítica. Isto é possível, visto existir uma solução em forma fechada, pois existe um número finito de funções conhecidas que definem o nosso modelo.

4. Questão

Com a formulação anterior, estão reunidas as condições para formular questões ao nosso modelo. A questão global é respondida através da distribuição à posterior do modelo, $P(c_n|p_j, f_i, \theta)$. Para além da questão global, o modelo permite responder também a questões intermédias, através de sub-questões paramétricas, visto que a formulação suporta o acesso a níveis singulares de informação. Este tipo de formulação é permitido, pois as variáveis de característica são consideradas independentes e identicamente distribuídas. Em suma, é permitido saber o estado de Laban dado o conhecimento de apenas uma determinada característica, tal como se pode verificar nos seguintes exemplos: $P(c_n|p_j), P(c_n|f_i)$ e $P(c_n|\theta)$. A Figura 4.4 mostra o nosso modelo completo.

Programa
- Descrição
  - Especificação
    - **Variáveis relevantes:**
    - $c_n \in \mathcal{L}$: representa o $n^{\text{ésimo}}$ componente Laban;
    - $p_j \in \hat{P}$: representa a $j^{\text{ésimo}}$ característica estática da silhueta;
    - $f_i \in F$: a $i^{\text{ésima}}$ característica dinâmica da silhueta;
    - $\theta \in [-\pi, \pi]$: representa o deslocamento relaticvo à silhueta;
    - **Decomposição:**
    - $P(c_n, p_j, f_i, \theta) = P(c_n)P(p_j|c_n)P(f_i|c_n)P(\theta|c_n)$
    - **Formulação:**
    - $P(c_n)$: uniforme;
    - $P(p_j, f_i, \theta|c_n)$: Distribuições gaussianas independentes;
  - **Identificação:**
  - nenhuma;
- **Questões:**
- $P(c_n|f_i, p_j, \theta) : P(c_n|f_i), P(c_n|p_j), P(c_n|\theta)$

Figura 4.4: Programa Bayesiano, que vai permitir estimar o estado mais provável cada uma das componentes de Laban c_n.

4.2.3 Aprendizagem

As distribuições de probabilidade, genericamente $P(\theta, p_j, f_i|c_n)$, representam o modelo de Laban, e são construídas a partir de conjuntos de dados de treino contendo sequências de movimento. O primeiro passo no processo de aprendizagem é a anotação manual da sequência de imagens que será usado para treinar o modelo. Considere uma sequência $\mathbf{I}$ anotada com um conjunto de estados Laban dominantes, um para cada componente c_n correspondente. Para cada $\hat{\mathbf{I}} \in \mathbf{I}$ calculamos um conjunto de características $\{p_j, f_i, \theta\}$. Para cada componente c_n, agrupam-se as características diferentes em relação a cada estado possível, ou seja, dois conjuntos de características são associados a classes q_1 e q_2. Por isso, considerando as características como independentes e igualmente distribuídas para um dado estado p_k, como resultado do processo associativo, temos uma distribuição de probabilidade definida como

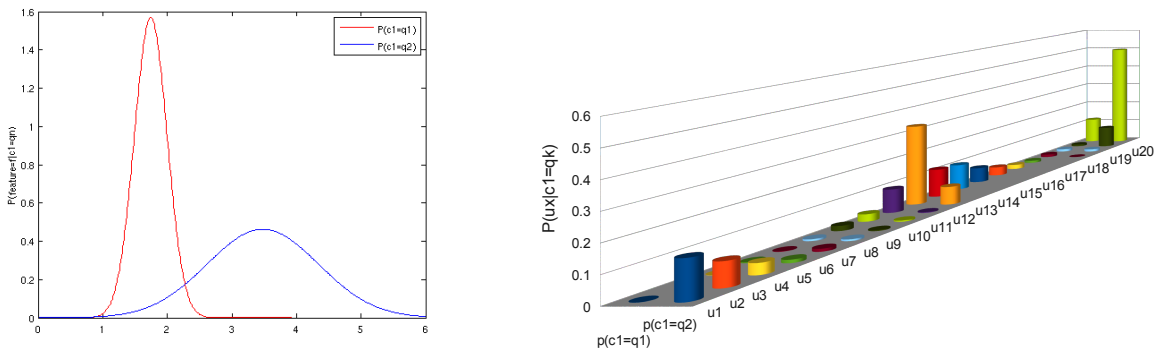
$$P(\beta|c_n = q_k) = \mathcal{N}(\mu_\beta, \sigma_\beta), \ \beta \in \{f_i, \theta\} \tag{4.4}$$

onde μ_β e σ_β representam a média e variância do conjunto de amostras para uma dada variável aleatória β.

No caso específico de p_j, conforme definido no nosso programa Bayesiano, esta variável é discretizada em x intervalos equidistantes, entre os valores mínimo e máximo observados para p_j, em todos os dados experimentais disponíveis. Assume-se que em caso de se observarem estados que ultrapassem os limites observados nos dados de treino, que se escolhe o intervalo na extremidade mais próxima. Cada intervalo corresponde a um único estado possível u, tal que, $p_j = \{u_1, \cdots, u_x\}$. Por isso, a formulação da probabilidade usando variáveis discretas, pode ser representada como uma matriz estocástica $M^{c_n}_{x,r}$, onde cada probabilidade da célula é dada pelo rácio entre o número de amostras para um dado u_x, normalizados pelo total.

$$P([p_j = u_x]|c_n = q_k) = \frac{\sum \text{de observações } u_x \text{ para o estado } q_k}{\sum \text{total de observações } u_x \text{ para o estado } q_k} \tag{4.5}$$

Na Figura 4.5 apresenta-se um exemplo da distrubuição resultante para cada uma das a-



(a) Distribuição Gaussiana (b) Histograma

Figura 4.5: Dois exemplos das formas paramétricas aprendidas, um para cada tipo, Gaussiana e Histograma.

proximações. A distruição Gaussiana (Figura 4.5a) é um exemplo calculado a partir da carácterística c_6, em que a distruição desenhada a vermelho representa $P(p_1|c_6 = q_1)$, sendo que a densidade a cor azul representa $P(p_1|c_6 = q_2)$. Como se pode constatar, a variável p_1 apresenta curvas discriminantes para os diferentes estados de c_6. Na Figura 4.5b pode ver-se um exemplo para c_4, em que a variável é discretizada em 20 estados diferentes, isto é, em diferentes classes u_x com $x = 20$. Analogamente ao verificado para a distruição Gaussiana, existem 2 distribuições, uma para cada estado $\{q_1, q_2\}$, cujas probabilidades são apresentadas para cada um dos estados. A soma das probabilidades ao longo de u_x é naturalmente 1.

4.2.4 Resultados Experimentais

As duas primeiras linhas da Tabela 4.1 apresentam a percentagem de estados correctamente identificados por componente, quando comparados com os estados anotados para cada instante de tempo, bem como a confiança média das estimativas do modelo, i.e. a probabilidade média com que os estados são classificados. A percentagem média *r.p.f.* é de 87.67%, ao passo que a confiança média *c.m.m.* é superior a 90%. Contudo, para a componente c_8, a precisão é baixa, apresentando confianças elevadas, o que significa que o modelo converge muitas vezes para um falso positivo, nesta característica particular. Isolando a componente c_8 das restantes, os resultados revelam um modelo preciso quando comparando as anotações e estimativas, mostrando que características visuais podem ser aplicadas com robustez na caracterização de sequências de movimento usando descritores semânticos de Laban. A análise de resultados é estendida com a Tabela 4.1, onde são apresentadas as características dominantes para cada actividade analisada, resultantes do processo de classificação. Os movimentos são normalmente caracterizados por propriedades específicas que, através do modelo proposto, são representados em conformidade por símbolos dominantes, e.g. acções tradicionalmente associadas a movimentos rápidos são associadas ao símbolo *sudden*.

A Figura 4.6 apresenta um exemplo da segmentação criada a partir do modelo de análise Laban baseado, onde é classificado o estado mais provável para cada componente. A representação topológica surge, escolhendo-se o estado mais provável, resultante da probabilidade à posteriori do modelo Laban, sendo que se as probabilidades para ambos os estados estiverem no intervalo $[0.4, 0.6]$, o estado é considerado indefinido. Analisando em detalhe o gesto aplicado da figura, podemos constatar a primeira componente Laban alterna de forma constante, sendo o movimento predominantemente repentino. Em alguns *frames* o movimento é classificado como *sustained*, derivado à paragem do seu movimento para inversão do sentido. Para a componente *Esforço Espaço* e *Esforço Fluxo*, a classificação converge quase por completo para os estados *indirect* e *free*, respectivamente. É fácil entender isso, já que os movimentos praticados pelos braços exercem um movimento pouco controlado e de uma forma não linear. Para a componente *Forma Direcional*, esta comuta mais entre os dois estados, pois o movimento é recriado de uma forma mais circular como também em forma de arco. Na componente *Forma Forma*, esta adquire maioritariamente o estado *Ball*

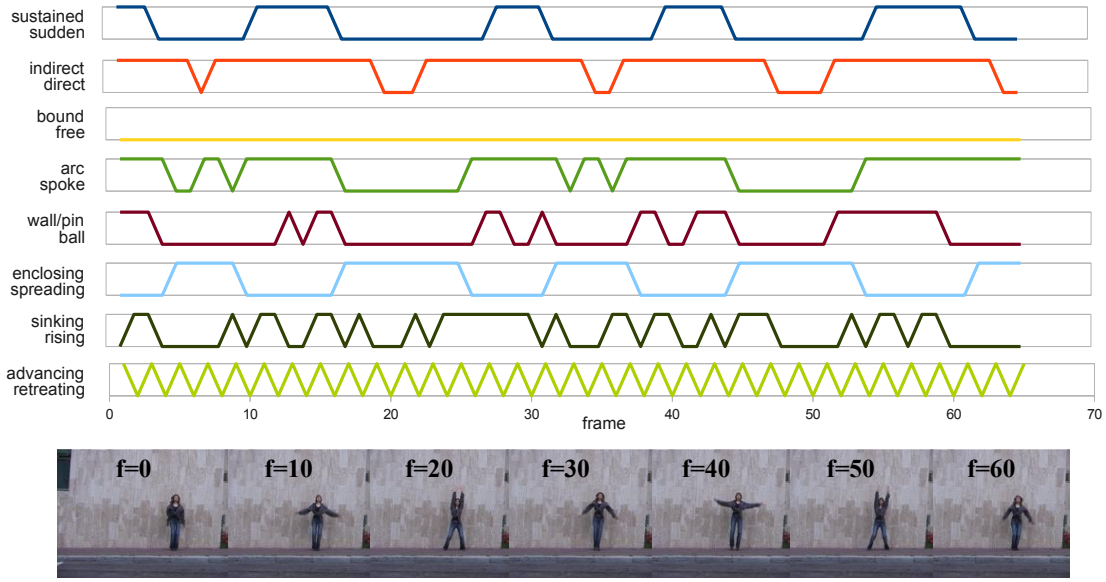
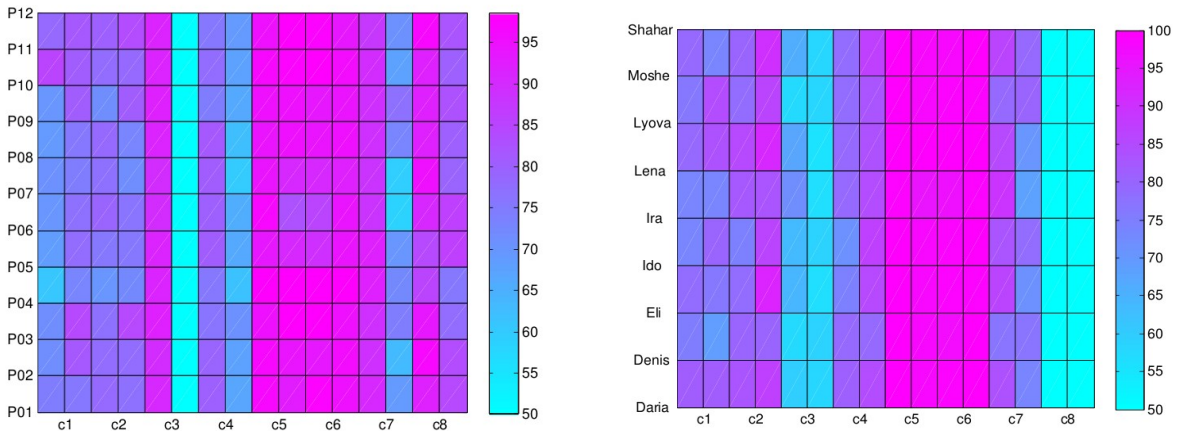


Figura 4.6: Exemplo da classificação simbólica Laban ($q_1 \oplus q_2$) para o gesto *saltar com palmas*, interpretado por *Daria* do dataset Weizmann em alguns *frames* chave.

já que na maioria dos *frames*, a intérprete tem as pernas e braços abertos. Para os últimos componentes da componente *Forma*, relativos à informação dimensional, é de salientar a *Forma X*, onde comuta de estado consoante o abrir e fechar braços e pernas. Para a *Forma Y*, a alteração de estado é muito acentuada, já que para abrir e fechar pernas, a intérprete aplica pequenos saltos. Estes pequenos e rápidos saltos, fazem o corpo em si subir/descer ao longo do gesto recriado. A última componente está representada em forma dente de serra, já que não foi possível definir um estado dominante durante a sequência.

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
r.p.f.%	94.02	83.17	91.43	67.02	80.34	81.91	63.17	21.43
c.m.m.%	87.42	93.18	90.58	82.88	89.24	98.91	90.46	90.88
bend	sustained	indirect	free	arc	n.d.	n.d	sinking	n.d.
jack	n.d.	n.d.	free	n.d.	n.d.	n.d.	rising	n.d.
jump	sudden	direct	free	spoke	pin	n.d.	rising	advancing
pjump	sudden	direct	free	spoke	pin	n.d.	rising	n.d.
run	sudden	direct	free	spoke	pin	enclosing	rising	n.d.
side	sudden	direct	free	spoke	pin	spreading	n.d.	n.d.
skip	sudden	direct	free	spoke	pin	n.d.	rising	n.d.
walk	sustained	direct	free	spoke	pin	n.d.	rising	advancing
wave1	sustained	indirect	free	arc	wall	spreading	n.d.	n.d.
wave2	sustained	indirect	free	arc	wall	spreading	n.d.	n.d.
boxing	sudden	direct	bound	spoke	nd	spreading	n.d.	advancing
handclapping	sustained	indirect	bound	bound	n.d.	n.d.	n.d.	n.d.
handwaving	sustained	indirect	free	n.d.	ball	spreading	rising	n.d.
jogging	n.d.	n.d.	free	spoke	n.d.	n.d.	n.d.	n.d.
running	n.d.	n.d.	free	spoke	n.d.	n.d.	n.d.	advancing
walking	n.d.	n.d.	free	spoke	pin	enclosing	n.d.	n.d.

Tabela 4.1: Estados de Laban dominantemente classificados. Cada estado é considerado dominante, se e só se for classificado com uma frequência de pelo menos $^2/_3$ do total de *frames* numa sequência de movimento dada por I. Um estado é considerado definido se a sua probabilidade estimada $P(c_n|f_i, p_j, \theta) \geq 0.4$. Os acrónimos *r.p.f* e *c.m.m* significam Rácio de Precisão por *Frame* e Confiança Média do Modelo respectivamente. O acrónimo *n.d.* significa estado Não-Dominante.



(a) KTH (b) WZ

Figura 4.7: Valor médio da estimação do classificador de Laban, para as diversas componentes (eixo das abcissas) relativamente aos diferentes actores (eixo das ordenadas), nas bases de dados consideradas. A escala de cores, representa o valor médio para a probabilidade de cada um dos estados das componentes Laban $P(c_n)$.

A análise aos resultados experimentais do modelo apresentado neste capítulo conclui-se com a apresentação de dois gráficos ilustrando a confiança média das estimativas geradas pelo classificador, em função das várias pessoas e diferentes estados para as componentes modeladas. Os gráficos apresentados na Figura 4.7 apresentam no eixo das ordenadas os diferentes actores presentes em cada uma das base de dados. O eixo das abcissas enumera cada uma das componentes c_n em que as colunas se encontram agrupadas aos pares, representando os estados q_1 e q_2 respectivamente. Na Figura 4.7a, podemos verificar (ao analisarmos cada coluna) que existem diferenças visíveis entre os valores médios de confiança para as diferentes pessoas em cada uma das componentes modeladas. Ainda assim, e apesar das diferenças, é possível identificar, através da análise de cada uma das linhas, que existem pessoas em que a confiança média é semelhante para a maioria os estados.

Conclusões semelhantes se podem tirar da análise ao gráfico da Figura 4.7b. Neste, é mais visível a homegeneidade com que as componentes são classificadas para os diferentes actores. Contudo, são visíveis algumas diferenças, ainda que seja possível observar pequenos grupos de pessoas onde os valores médios ao longo das abcissas são, de facto semelhantes.

Da análise destes resultados, conclui-se que, existe precisão e diferenças de confiança média suficientes, que motivam o desenvolvimento de um sistema para identificar diferentes pessoas. Este sistema pretende explorar as diferenças mencionadas, para que seja possível identificar uma pessoa, com base nas características de movimento específicas a cada um dos actores.

Capítulo 5

Modelo Bayesiano para Identificação de pessoas

No trabalho de Santos e Dias demonstram que a descrição de movimentos usando Análise de Movimento de Laban pode ser generalizada simbólicamente [Santos and Dias, a]. Na sua investigação, bem como neste trabalho, os estados Laban mostram ser repetitivos para acções semelhantes, mesmo quando praticadas por diferentes pessoas. Contudo, a confiança com que o modelo classifica cada estado, difere de pessoa para pessoa, como se verifica na análise aos resultados da secção anterior. Esta propriedade indica que o espaço Laban pode ser discriminante em relação a quem está a executar uma sequência observável. O modelo visual Laban apresentado é uma solução modificada em relação aos modelos Laban baseados em trajectórias [Santos and Dias, a] e do sistema de identificação de pessoas [Santos and Dias, b].

5.1 Modelos de Assinatura e Identificação

Definição Espaço Laban [Santos and Dias, b]: Considere o Espaço Laban $\chi \in \mathbb{R}^n$ como sendo uma representação de dimensão n, para cada uma das variáveis Laban c_n. Assuma o vector $R \in \chi : R = (\tau_1, \cdots, \tau_n)$, que representa uma combinação das propriedades Laban para a sequência de um movimento, onde $\tau_n = P(c_n = q_1)$.

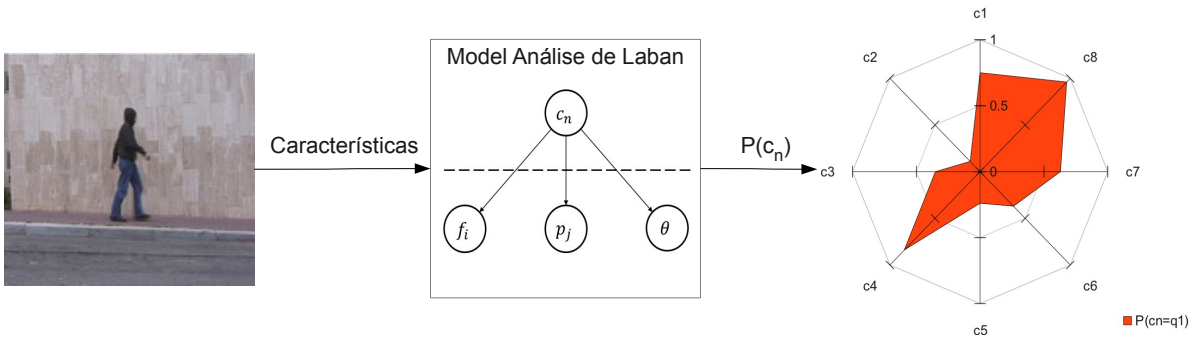


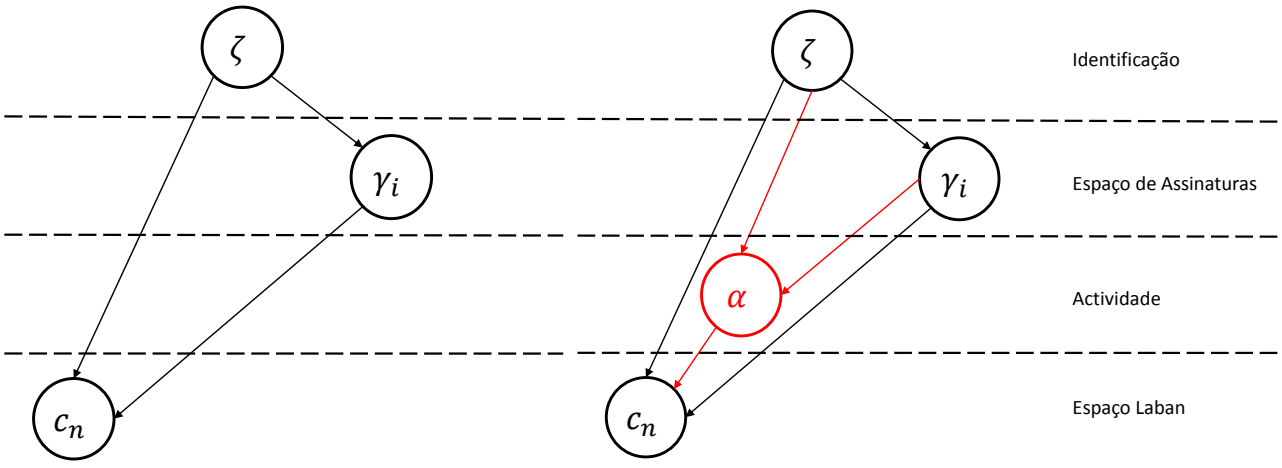
Figura 5.1: Representação gráfica do Espaço de Laban, onde se pretende mostrar que este resulta do espaço probabilístico resultante do modelo de classificação de Laban.

Esta simplificação, é feita devido aos estados mutuamente exclusivos, evitando assim redundância, pois o espaço de Laban para q_2 iria conter as probabilidades complementares

de q_1. A Figura 5.1 ilustra o Espaço Laban verificado para o instante do frame de movimento (lado esquerdo da imagem) também apresentado. Na representação gráfica podem observar-se tantos eixos quantas as componentes de Laban, sendo que a forma geométrica (a cor-de-laranja) surge após fechar o contorno sobre os pontos em cada eixo.

Neste trabalho, vai-se usar a aproximação de assinatura Laban que melhor performance demonstrou em [Santos and Dias, b]. Segue-se a sua descrição. Considerando cada variável do espaço Laban como um nó topológico, são medidas as distâncias entre nós como sendo $d_{i,j} = \tau_i - \tau_j$. A assinatura é posteriormente criada definindo a matriz adjacente A do gráfico topológico, a partir da qual se calculam os parâmetros da sua decomposição em valores singulares. As variáveis assinatura $\Gamma = \{\gamma_i\}$ são independentes e definidas a partir dos parâmetros atrás referidos, mais concretamente os valores e vectores próprios calculados.

Modelos de assinatura e identificação: Em [Santos and Dias, b], o Laban é aplicado para analisar diferentes partes do corpo ao invés do corpo como um todo, que naturalmente aumenta a capacidade discriminante da assinatura. Experiências iniciais (a serem apresentadas na próxima secção 5.2), usando o modelo na sua versão original, **Modelo A** (5.2a), mostraram uma fraca precisão no reconhecimento de pessoas. A análise detalhada dos resultados mostraram que a generalização o Laban para o corpo como um todo, conduz a estimativas de confianças semelhantes no modelo Laban. Isto significa que para diferentes grupos pessoas, os estados Laban são estimados com valores de probabilidade semelhantes, desde que nas mesmas categorias de acções, o que causa uma elevada taxa de identidades incorrectamente classificadas. Para contornar esse problema, propomos uma versão modificada, onde se vai adicionar uma variável extra ao modelo, a actividade $\alpha \in \Lambda$, tal como ilustrado na Figura 5.2b, resultando na aproximação denominada **Modelo B**.



(a) Modelo A (b) Modelo B

Figura 5.2: Modelo gráficos (a) Original, e (b) Modificado

Considere $\gamma_i \in \Gamma$ uma variável de assinatura de movimento, independente e igualmente distribuída. Podemos definir a seguinte distribuição conjunta para o modelo de identificação:

$$P(\zeta, \alpha, \gamma_i, c_n) \tag{5.1}$$

onde ζ representa uma variável de reconhecimento, cujos estados correspondem a diferentes identidades. Auxiliados pela Figura 5.2b obtém a decomposição dada pela equação (5.2).

$$P(c_n|\alpha)P(c_n, \alpha|\gamma_i)P(\alpha, \gamma_i, c_n|\zeta)P(\zeta) \tag{5.2}$$

Na atribuição de formas paramétricas, a distribuição à priori $P(\zeta)$ é inicializada como sendo uma distribuição uniforme. A actividade Λ é estimada baseada nas combinações observadas das variáveis Laban, para a qual é válida a seguinte decomposição, $P(c_n|\alpha)$. Na nossa aproximação modificada, o modelo de assinatura é aprendido como sendo *kernel* de distribuições $P(\gamma_i|c_n, \alpha)$, onde a variável α aparece como indexador de distribuições para as diferentes actividades. Quer-se com isto dizer que para cada actividade $\alpha \in \Lambda$, é calculada uma distribuição de assinatura para as diferentes variáveis $\gamma_i \in \Gamma$, processo que é válido para todas as diferentes pessoas presentes nos dados de treino. Usando a regra de Bayes e a regra da marginalização, a decomposição resultante é descrita na equação seguinte,

$$P(\zeta|\alpha, \gamma_i, c_n) \propto P(\zeta) \prod_{q=1}^{n} P(c_q|\alpha) \prod_{q=1}^{n} P(\alpha, c_q|\gamma_i) \prod_{q=1}^{n} \prod_{p=1}^{i} P(c_n, \gamma_p, \alpha|\zeta) \tag{5.3}$$

onde o factor normalização é omitido por motivos de simplificação. As distribuições $P(c_n|\alpha)$, $P(c_n|\gamma_i)$ e $P(\alpha, \gamma_i, c_n|\zeta)$ são as distribuições de probabilidade que constituem o *corpo* do modelo de identificação, e são treinadas a partir de dados experimentais previamente disponíveis. As distribuições $P(c_n|\alpha)$ são gaussianas, enquanto que $P(\alpha, c_n|\gamma_i)$ é um *kernel* de distribuições gaussianas para γ_i gerados a partir dos valores de probabilidade de c_n e indexados por α. A ultima distribuição é representada por uma matriz estocástica multivariável onde as assinaturas são associadas a identidades por meios de actividades e indexação de estados de Laban. Como se é fácil constatar em relação 5.2a, a complexidade do modelo aumentou, consequência da existência de uma nova variável para as acções e das novas depêndencias resultantes.

5.2 Resultados Experimentais

O nosso setup experimental para a identificação envolve duas bases de dados conhecidas na área de análise de movimento, KTH[1] e WZ[2]. Cada sequência de movimento foi analisada em termos dos descritores simbólicos Laban, como anteriormente apresentado, que são posteriormente usados para gerar assinaturas de movimento e classificação de identidades. Nesta secção, temos interesse em conhecer a taxa de identificação, i.e. quantas vezes uma pessoa é correctamente identificada. A classificação é apresentada por frame e por sequência.

[1][Online] `http://www.nada.kth.se/cvap/actions/`
[2][Online] `http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html`

5.2.1 Dataset KTH

O **dataset KTH** (Figura 5.3) contém 6 acções diferentes (andar, *jogging*, correr, socar, acenar a mão,bater palmas) realizado por 25 pessoas diferentes, das quais neste trabalho foram escolhidas aleatoriamente 12, enumeradas de p01 a p12. As sequências de vídeos são adquiridas à taxa de 25 *fps* frames e resolução 160×120, com 4 segundos de duração média por tentativa executada. Chama-se à atenção do leitor que cada vídeo disponível contém várias execuções da mesma pessoa.

Figura 5.3: Diferentes tipos de acções presentes no dataset do KTH.

5.2.2 Dataset Weizmann (WZ)

O **dataset Weizmann** (Figura 5.4) é um conjunto de sequências de 90 vídeos de baixa resolução 180×144), gravados a uma taxa de 50 *fps* frames. Tem 9 pessoas diferentes, realizando 10 acções diferentes (Correr, Andar, *Skip*, Saltar com Palmas, Saltar para a Frente, Saltar (no mesmo lugar), Correr Lateralmente, Acenar com uma e duas mãos, Baixar(-se)).

Figura 5.4: Diferentes tipos de acções presentes no dataset do Dataset Weizmann

5.2.3 Métricas de Análise de Resultados

Para a análise dos resultados, faz-se uso de matrizes de confusão, bem como de duas métricas bastante populares no estudo de algoritmos de classificação. Chama-se a atenção do leitor que se irá fazer uso dos termos em inglês *precision* and *recall*. *Precision* é uma métrica usada para medir o número de amostras correctamente identificadas num conjunto de amostras pertencentes todas à mesma classe (geralmente medida em cada **linha** da matriz de confusão), e é dada por:

$$precision = \frac{verdadeiro\ positivo}{verdadeiro\ positivo + falso\ positivo}. \tag{5.4}$$

Esta métrica é geralmente usada em conjunto com o *recall*, que representa o rácio de classificações relevantes de entre todas as classificações, cujas estimativas apontam para a mesma classe (geralmente medida em cada **coluna** da matriz de confusão), tal que:

$$recall = \frac{verdadeiro\ positivo}{verdadeiro\ positivo + falso\ negativo}. \tag{5.5}$$

Um verdadeiro positivo verifica-se quando a classes original e estimada são iguais. O falso positivo para uma classe C verifica-se quando o analisar uma qualquer classe $\neg C$, o resultado estimado é C. O falso negativo para a classe C é observado quando a estimativa para uma sequência pertencente classe C é classificado como sendo pertencente a outra $\neg C$.

5.2.4 Resultados Para o Modelo A

	p01	**p02**	**p03**	**p04**	**p05**	**p06**	**p07**	**p08**	**p09**	**p10**	**p11**	**p12**
p01	118	24	0	2154	47	0	24	0	0	0	0	0
p02	0	1568	0	304	117	70	23	0	94	47	117	0
p03	0	286	0	0	94	0	7	0	114	143	1916	249
p04	0	0	24	2050	283	0	0	0	0	0	0	0
p05	0	0	0	1068	285	0	332	356	0	332	0	0
p06	65	0	0	2089	0	0	0	0	0	0	0	0
p07	0	103	0	0	0	672	0	0	362	0	1448	0
p08	181	0	0	2304	0	0	0	104	0	0	0	0
p09	0	0	131	157	0	0	0	209	0	0	2115	0
p10	0	348	75	124	50	0	0	0	0	0	1891	0
p11	0	0	0	0	0	0	0	0	0	25	2467	0
p12	0	11	0	0	0	0	0	0	0	49	2960	60

Tabela 5.1: Resultados para a identificação no Dataset do KTH para o modelo de identificação original: Tabela de confusão com a precisão da classificação por frame.

Lembra-se o leitor, que a esta aproximação, corresponde o modelo gráfico da Figura 5.2a, ou seja, sem que as diferentes actividades sejam tomadas em conta. Como pode ser verificado pelas matrizes de confusão nas Tabelas 5.1 e 5.2, a precisão ao longo do tempo (por frame) é extremamente baixa. Este resultado naturalmente reflecte-se na precisão final, onde a performance do sistema atinge o seu pico mais baixo. Estes resultados podem ter a sua justificação baseada nas tabelas de classificação dos estados Laban por diferentes pessoas, da secção anterior. O facto de se generalizarem os descritores de Laban para o corpo

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Daria	3	0	7	13	0	0	0	0	438
(2) Denis	1	297	4	7	13	1	17	0	1
(3) Eli	0	12	16	0	1	12	0	0	328
(4) Ido	0	0	3	0	2	0	0	243	0
(5) Ira	2	0	0	0	0	4	0	0	580
(6) Lena	0	0	0	0	0	0	0	8	407
(7) Lyova	0	0	1	0	0	31	0	0	215
(8) Moshe	0	0	0	0	0	15	0	333	3
(9) Shahar	0	0	0	7	5	2	0	0	340

Tabela 5.2: Resultados para a identificação no Dataset do Weizmann para o modelo de identificação original: Tabela de confusão com a precisão da classificação por frame.

como um todo, não permite uma descrepância nos valores absolutos da estimação, que seja suficiente discriminativa, de tal modo a que esta seja invariante à actividade realizada. Desta interpretação colocou-se a hipótese de incluir uma variável extra no modelo de identificação, como descrito na sub-secção anterior.

5.2.5 Resultados Para o Modelo B

Reconhecimento de actividades nas bases de dados KTH e Weizmann

Nesta secção, apresentam-se os resultados da classificação de actividades com base na questão Bayesiana $P(\alpha|c_n)$ proposta no Modelo B. Conforme visível no gráfico da Figura 5.5, ambas as bases de dados apresentam alta precisão, o que se irá reflectir na identificação das diferentes pessoas. De notar que se agruparam as actividades comuns a ambas as bases de dados. A precisão é menor nos gestos *Salto com Palmas* que é confundido algumas

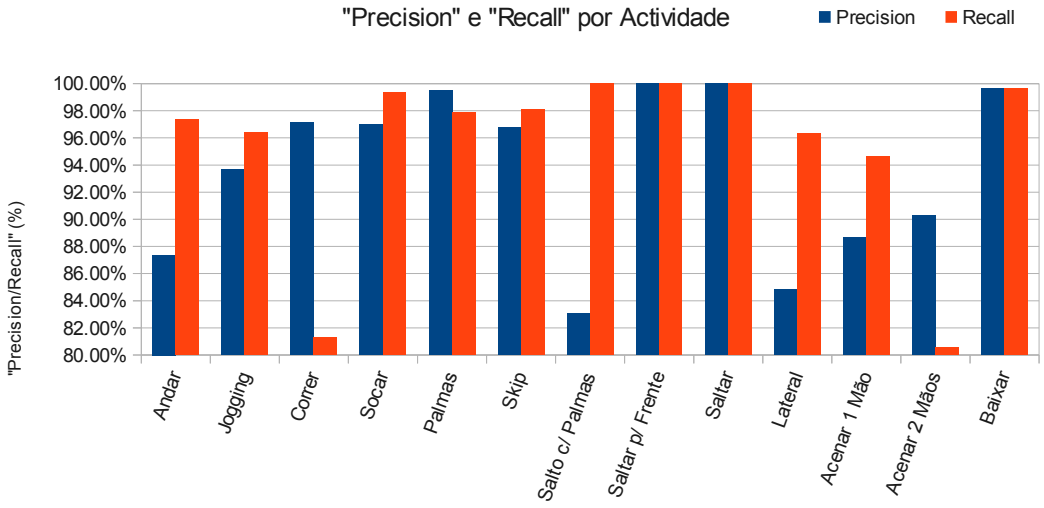


Figura 5.5: Valores do Precision e Recall por frame para as várias acções presentes nas bases de dados KTH e Weizmann.

vezes com o *Acenar 2 Mãos*, bem como a actividade *Correr Lateralmente* que se confunde ocasionalmente com as actividades *Correr* e *Jogging*. Este decréscimo de precisão deve-se ao facto de estas actividades produzirem por vezes formas de silhueta semelhantes, o que se propaga na classificação das componentes Laban e, por seu turno, na actividade.

Contudo, a precisão conseguida é de vital importância para a nossa adaptação do modelo de identificação, visto que este tem como uma das variáveis centrais, a variável α de actividade. Simultaneamente o nosso modelo demonstra além de uma boa capacidade de análise de movimento, uma capacidade para conseguir discriminar entre um conjunto de actividades diferentes.

Identificação de pessoas na base de dados KTH

Os resultados da identificação para o dataset do KTH estão incluídos na Tabela 5.3. Analisando os valores absolutos por frame, verificamos rapidamente que os resultados são bastantes satisfatórios, apesar de em algumas sequências a confusão ser maior. Exemplos disso são os sujeitos p04, p06 e p09 que exibem um alto número de falsos negativos, ou seja, as suas identidades são identificadas como sendo correctas, durante a execução por parte de actores diferentes. Na Figura 5.6, e para o mesmo sujeito percebe-se o porquê de a probabilidade do *recall* por frame ser mais baixa relativamente aos restantes sujeitos, o que ainda assim não invalida a elevada precisão na identificação, para a globalidade dos executantes, sendo que apenas o actor p01 exibe precisão abaixo dos 90%. Os resultados apresentados

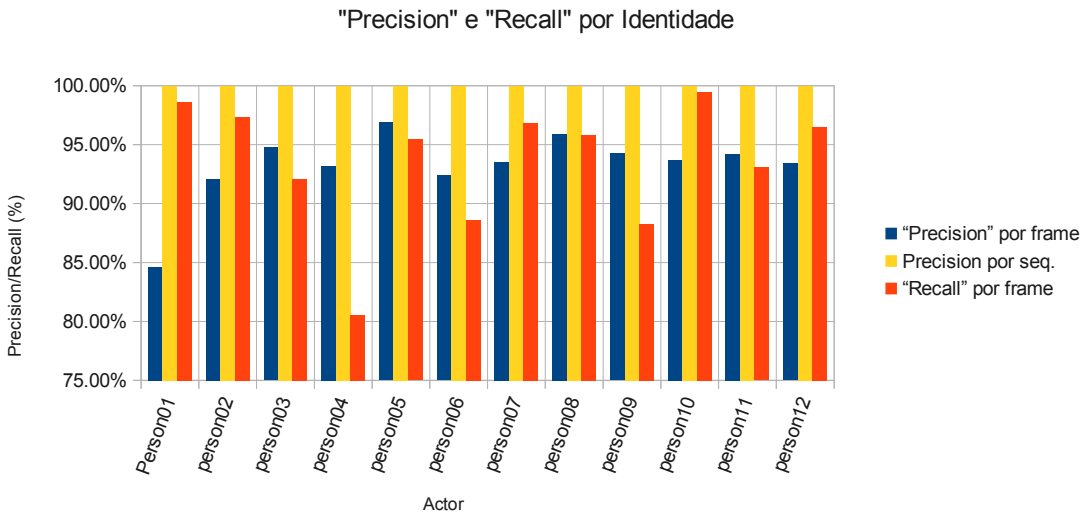
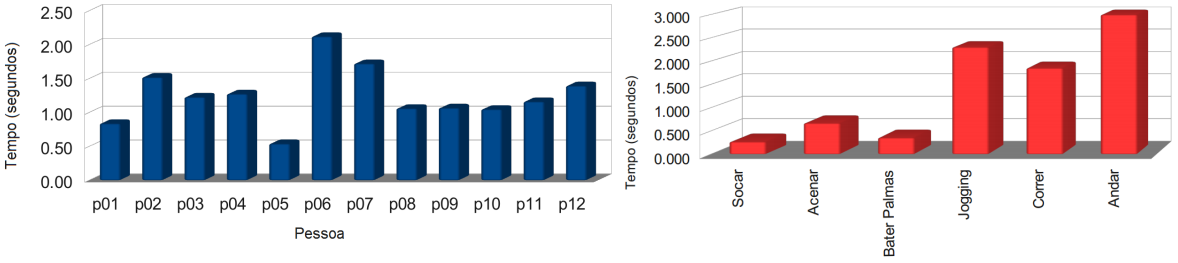


Figura 5.6: Valores do Precision e Recall de cada sujeito para a base de dados KTH

	p01	**p02**	**p03**	**p04**	**p05**	**p06**	**p07**	**p08**	**p09**	**p10**	**p11**	**p12**
p01	2003	15	2	67	0	172	0	41	4	0	59	4
p02	0	2156	72	1	0	0	3	2	24	0	77	6
p03	6	1	2710	43	17	0	0	1	53	8	7	14
p04	1	0	12	2196	5	0	0	0	99	0	16	27
p05	0	0	1	38	2300	1	0	28	4	1	0	1
p06	10	10	46	25	10	1990	52	1	7	0	3	0
p07	2	12	15	53	0	80	2419	0	5	0	0	0
p08	2	0	7	26	35	0	0	2482	30	0	7	0
p09	0	4	1	138	2	0	0	0	2461	3	2	0
p10	0	0	9	22	41	0	0	17	16	2331	4	48
p11	0	17	62	24	0	3	0	0	35	0	2348	3
p12	7	0	5	93	0	0	24	19	50	1	0	2821

Tabela 5.3: Resultados para a identificação na base de dados do KTH: Tabela de confusão da classificação por frame; (precisão global por sequência = 100%).

referem-se à análise de cada frame de todas as sequências disponíveis. Contudo, uma medida importante é também aferir-se qual a identidade estimada no fim de cada sequência. Neste campo, os resultados são plenamente satistórios, onde a taxa de sucesso é de 100% para todas as pessoas analisadas, conforme ilustra a recta amarela do gráfico acima mencionado. Complementam-se os resultados para esta base de dados com informação adicional em termos da *velocidade* de convergência. A Figura 5.7a mostra o instante tempo em que a estimativa converge para a decisão final acerca do actor, cujo movimento é observado. Esta medida tem em conta o último *frame* a partir do qual todos os restantes são identificados como sendo a identidade correcta. A convergência acontece em média entre o instante $t = 1s$ e $t = 2s$, demonstrando assim que a convergência é atingida após a observação de curtas sequências, satisfazendo assim o requisito de tempo real. Assume-se que tempos inferiores a $2s$ são suficientes para satisfazer as necessidades de um sistema inteligente em tempo real. São também mostrados os valores de convergência, mas desta forma seleccionados pelas diferentes acções na Figura 5.7b. Neste caso, as situações onde o tempo de convergência é maior, verificam-se simplesmente pelo facto de algumas acções serem muito semelhantes,i.e., andar, correr ou *jogging*.



(a) Por Pessoa.

(b) Por Acção.

Figura 5.7: Tempo médio de convergência , (a) por pessoa (b) por acção, considerando sequências de video com fps = 25 *frames* por segundo.

Identificação de pessoas na base de dados Weizmann

Relativamente ao dataset do *Weizmann Institute*, os resultados da identificação estão expressos na matriz de confusão da Tabela 5.4. Os valores absolutos resultantes por *frame* para este dataset também são promissores. À semelhança do dataset do KTH, algumas

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Daria	379	9	1	26	3	4	26	10	3
(2) Denis	26	267	0	16	1	26	2	3	0
(3) Eli	9	0	341	5	3	1	7	2	1
(4) Ido	25	0	1	198	3	6	4	11	0
(5) Ira	2	5	10	12	491	16	12	35	3
(6) Lena	22	5	0	0	6	376	5	1	0
(7) Lyova	2	1	0	7	0	20	212	5	0
(8) Moshe	0	0	3	1	3	9	1	334	0
(9) Shahar	7	1	3	4	4	0	23	4	308

Tabela 5.4: Resultados para a identificação na base de dados do Weizmann: Tabela de confusão da classificação por frame; (precisão global por sequência = 100%).

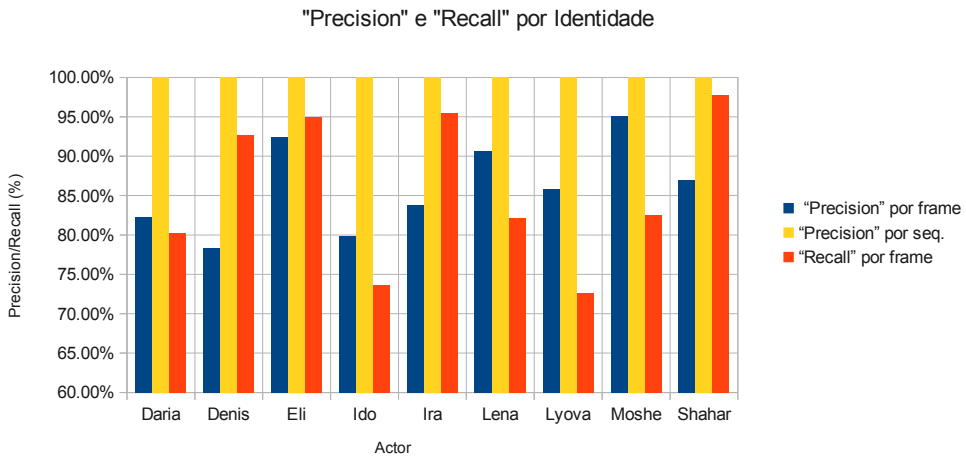
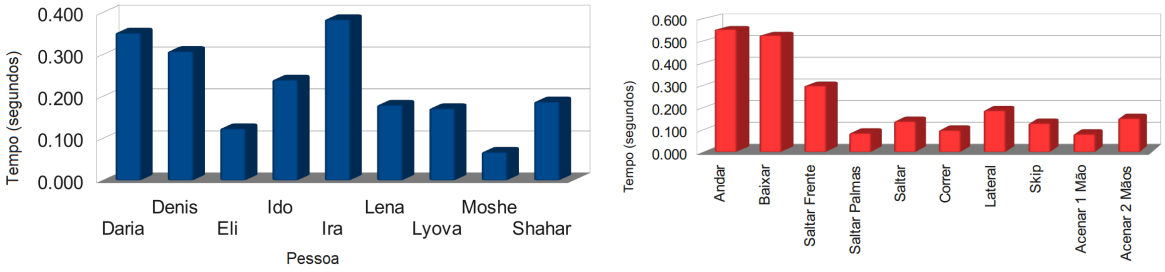


Figura 5.8: Valores do Precision e Recall de cada sujeito para o Dataset KTH



(a) Por Pessoa.

(b) Por Acção.

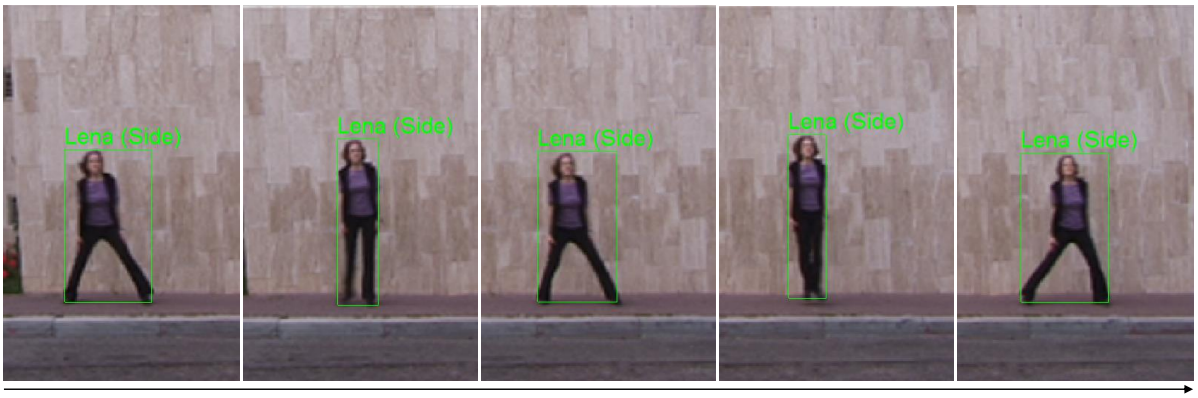
Figura 5.9: Tempo médio de convergência , (a) por pessoa (b) por acção, considerando sequências de video com fps = 50 *frames* por segundo.

amostras apresentam uma maior confusão não sendo crítico em nenhum dos casos analisados. Os maiores números de falsos positivos são observados para a *Daria*, *Ido* e *Lyova*, sendo que em termos de falsos positivos, os actores *Denis* e *Ido* são os actores onde estes mais se verificam. A Figura 5.8, mostra probabilidades do *precision* e *recall* demonstrando rácios ligeiramente inferiores ao dataset anterior. Especificamente na medida *recall*, as percentagens observadas exibem maior inconstância, sem contudo isto significar que uma performance média situada entre os 85% e os 90% seja considerado um mau resultado. Tal como no dataset anterior, a classificação por sequência atinge 100% de precisão. Este factor ganha maior relevância, se tivermos em conta a existência de várias acções que são semelhantes. O rácio por *frame* apresenta para este dataset um elevado número de precisões abaixo dos 90%, contudo continuamos a considerar que seja uma performance elevada. O facto da performance por sequência ter renovado a performance de 100%, reforça ainda mais o sucesso dos objectivos traçados na secção 1.1 do capítulo 1. A Figura 5.9 mostra o tempo por pessoa e por acção.

O tempo de convergência é mais rápido que no KTH, que é, em parte justificado por uma taxa *fps* de aquisição maior.

Conclui-se os resultados experimentais com um exemplo de como funciona o sistema de identificação de pessoas. Como se pode ver na Figura 5.10, a pessoa é detectada, as

(a) Exemplo para a acção *Andar* executada por *Daria*.



(b) Exemplo para a acção *Andar Lateralmente* executada por *Lena*.

Figura 5.10: Exemplos de classificação de identidade, sobrepostas sobre a imagem original, contendo também informação sobre a actividade.

características calculadas e a sua identificação sobreposta aos frames originais. O código necessário a correr este e outros exemplos está previsto ser disponibilizado em breve numa página de suporte a este trabalho `www.isr.uc.pt/~luis"`.

Capítulo 6
Conclusões e Trabalho Futuro

6.1 Sumário do Trabalho

Este trabalho pretendeu resolver dois problemas identificados no estado de arte, nas áreas de identificação de pessoas e desenvolvimento de modelos computacionais para Análise de Movimento Laban. Neste trabalho o estudo concentrou-se maioritariamente em duas grandes áreas de investigação: Visão por Computador e Reconhecimento de Padrões. Destacam-se duas grandes contribuições, conforme se apresenta de seguida.

- **Um modelo probabilístico para representação de movimento baseado na Análise de Movimento Laban e em característcas visuais.** Foram estudados métodos de representação alternativa a características visuais clássicas, por forma a se conseguir caracterizar o movimento nas suas componentes dinâmica e estática. Foi usado um classificador Bayesiano, usando um método chamado Programação Bayesiana, com o objectivo de avaliar o modelo probabilístico proposto, no que diz respeito às suas capacidades de análise do movimento. A Análise de Movimento Laban é a base de um sistema de notação capaz de analisar e caracterizar todo o tipo de movimento, definida em diversos componentes.

- **Um modelo de identificação de pessoas com base na descrição simbólica do seu movimento.** A classificação simbólica gerada do nosso classificador, foi integrada com um algoritmo de geração de assinaturas de movimento [Santos and Dias, b]. O objectivo inicial era a simples integração do modelo visual de Laban num sistema existente para identicar pessoas com base no seu movimento. Contudo, as simplificações consideradas no modelo de Laban, obrigam a que o modelo de identificação proposto em [Santos and Dias, b] tivesse de ser modificado. Nesse sentido, foi adicionada uma variável ao modelo existente, aumentando assim a sua complexidade. Apesar de se relaxar a independência da actividade, a precisão mostrada pelo modelo nos resultados mostrados nesta tese, abre grandes perspectivas para a continuação do seu desenvolvimento.

6.2 Conclusões

Neste trabalho foi desenvolvido um modelo para uma caracterização compreensiva de sequências de movimento, através do uso de características visuais. Este foi auxiliado por uma gramática descritiva adequada, baseada no principio da análise de Movimento Laban e na sua notação. A estrutura analisada foi integrada num sistema de identificação de pessoas previamente desenvolvido, onde os modelos foram adaptados para lidarem com a generalização Laban para o movimento corporal como um todo, em contraponto com a análise a partes específicas do corpo. O descritor simbólico Laban gerado foi aplicado na identificação de actividade para criar assinaturas de movimento, que ao ser combinadas mostraram boa capacidade discriminativa entre as diferentes pessoas. Para auxiliar todo o processo foi usando um classificador Bayesiano.

Os resultados mostraram ser promissores devido a elevada precisão e velocidade de identificação, dando indicações do potencial trabalho a ser desenvolvido no futuro para aperfeiçoamento e extensão do modelo proposto. O objectivo principal para este trabalho foi alcançado; um sistema de identificação de pessoas através da sua maneira de se mover. De realçar que o resultado deste trabalho foi, em grande parte, submetido e aceite sob forma de artigo de conferência (Anexo A).

6.3 Trabalho Futuro

O nosso trabalho é apenas um passo na direcção de combinações de actividades de pessoas e informação de movimento para melhorar os resultados obtidos na tarefa de identificação de pessoas. O problema aqui tratado não está de forma alguma encerrado/resolvido. O estado de arte sobre identificação de pessoas inclui soluções válidas para funcionar em cenários e com condições específicas/limitadas. Assim identificamos quatro áreas para um trabalho futuro:

- **Aumentar o modelo para várias câmaras simultaneamente.** Temos a intenção de explorar o comportamento das características em relação à aquisição de diferentes perspectivas aumentando o modelo para várias câmaras simultaneamente.
- **Melhorias no modelo de assinaturas.** É nossa expectativa melhorar o modelo de assinaturas no âmbito de relaxar a dependência de actividades específicas.
- **Melhorias na segmentação de imagem.** Pretendemos continuar a validar a nossa precisão na estimação de identidades com conjuntos de dados mais complexos, melhorando assim a segmentação da imagem.
- **Implementação prática.** Fornecer um protótipo funcional, para disseminação académica, a ser disponibilizado numa página web de suporte.

Bibliografia

[Benesh and Benesh, 1983] Benesh, R. and Benesh, J. (1983). Reading dance: The birth of choreology. In *McGraw-Hill Book Company Ltd.*

[Bessiere et al., 2012] Bessiere, P., Ahuactzin, J.-M., Mekhnacha, K., and Mazer, E. (2012). *Bayesian Programming.* Chapman and Hall/CRC.

[Eshkol and Wachmann, 1958] Eshkol, N. and Wachmann, A. (1958). Movement notation. In *Weidenfield and Nicholson.*

[Hamdoun et al., 2008] Hamdoun, O., Moutarde, F., Stanciulescu, B., and Steux, B. (2008). Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In *Distributed Smart Cameras, 2008. ICDSC 2008. Second ACM/IEEE International Conference on*, pages 1 –6.

[Han and Bhanu, 2006] Han, J. and Bhanu, B. (2006). Individual recognition using gait energy image. *TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 28(2):316–322.

[Held et al., 2012] Held, C., Krumm, J., Markel, P., and Schenke, R. (2012). Intelligent video surveillance. *Computer*, March:83–84.

[Iosifidis et al., 2012] Iosifidis, A., Tefas, A., and Pitas, I. (2012). Activity-based person identification using fuzzy representation and discriminant learning. *Transactions on Information Forensics and Security*, 7(2):530–542.

[Iwashita and Kurazume, 2009] Iwashita, Y. and Kurazume, R. (2009). Person identification from human walking sequences using affine moment invariants. In *International Conference on Robotics and Automation*, pages 436–441.

[Kendon, 2004] Kendon, A. (2004). Gesture: Visible action as utterance. Technical report, Cambridge University Press.

[Kouno et al., 2012] Kouno, D., Shimada, K., and Endo, T. (2012). Person identification using top-view image with depth information. In *13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, volume 1, page 6.

[Little and Marsh, 1992] Little, M. E. and Marsh, C. G. (1992). la danse noble, an inventory of dances and sources. In *Broude Brothers Ltd.*

[Murphy, 2002] Murphy, K. P. (2002). *Dynamic Bayesian Networks: Representation, Inference and Learning.* PhD thesis, University of California, Berkeley.

[Santos and Dias, a] Santos, L. and Dias, J. Laban-based multilayer model for activity recognition and annotation. (under peer review) on International Journal on Information Sciences, Elsevier.

[Santos and Dias, b] Santos, L. and Dias, J. Person identification based on activity invariant signatures. (under peer review) on Journal of Pattern Recognition Society, Elsevier.

[Yam et al., 2002] Yam, C., Nixon, M. S., and Carter, J. N. (2002). On the relationship of human walking and running: Automatic person identification by gait. In *16th International Conference on Pattern Recognition*, volume 1, page 4.

[Yu et al., 2006] Yu, S., Tan, D., and Tan, T. (2006). A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *The 18th International Conference on Pattern Recognition.*

Anexo A
Publicações

Artigo aceite para publicação no *2nd Workshop on Recognition and Action for Scene Understanding, REACTS 2013*

Vision-based Laban Motion Signatures for Person Identification

Luís Santos[1], José Sousa[1], and Jorge Dias[12]

[1] Institute of Systems and Robotics
Department of Electrical and Computer Engineering
University de Coimbra, Pinhal de Marrocos - Polo II
3030-290 Coimbra, Portugal
[2] ECE/Robotics at Khalifa, University (UAE), Abu Dhabi
{luis,jsousa,jorge}@isr.uc.pt
http://mrl.isr.uc.pt

Abstract. In this manuscript, a probabilistic model for *Visual* Laban Movement Analysis, based on human body silhouettes, is proposed as an extended solution to the problem of person identification. Laban models presented in literature are mostly based on features acquired from precise tracking technologies, whereas our method is suggested as a vision-based approach, generalizing the Laban model to visual-cues. Given the silhouette's dimension, we suggest an alternative representation based on Principal Component Analysis and the generalized Fast Fourier Transform. Laban symbolic descriptors are learned based on sets of training data, which are posteriorly used to characterize unknown sequences applying a Bayesian classifier. The proposed method has been successfully integrated with an adapted version of a person identification classifier, creating the grounds for developing an intelligent video-surveillance system. Results demonstrate an accurate classification model, in which features show to be discriminant to different performers, when performing the same actions. The experimental set-up shows capability for motion analysis, activity recognition and person identification, for which some demonstrations are publicly available.

1 Introduction

" *Video cameras are increasingly prevalent in society in both public and private spaces. At the same time, the quality of video surveillance continues to improve. This is especially true of intelligent video surveillance technology, which can recognize or track objects as well as identify human faces and behaviour patterns.*" - Held in [1]. Recently, Santos and Dias [2] developed a framework for activity invariant person identification. Their findings show an highly accurate framework which, using trajectory-based motion signatures, is capable of classifying between different person identities. This is a study of special interest in the area of video-surveillance. However, as they opportunely state, to make their framework straightforward applicable, there is an unsolved challenge to be addressed: associating image-based features to the set of activity invariant descriptors based on **Laban Movement Analysis (LMA)** [3].

1.1 Related work

Person Identification research can be broadly divided in two distinct categories: invasive and non-invasive biometrics. The majority of existent biometric works focus on fingerprint, iris or face analysis, which are *invasive* techniques, requiring some sort of cooperative behaviour by the identified person. Non-invasive researches address motion analysis in order to discriminate between different persons, focusing on gait as the primary activity, e.g. [4,5,6,7]. Kobayashi and Otsu [8], more recently Iosifidis et al. and Santos and Dias address activities other than gait, which present interesting results in the area of person recognition. Iosifidis et al.[9] address eating and drinking, while Santos and Dias [2] present an activity invariant solution, testing up to 9 different activities. An overview on the state of the art shows that action-based person recognition systems are still an open problem, which is increasingly receiving attention from the scientific community.
LMA is a motion notation language, developed in the context of dance and choreography by Rudolf Laban, which in the past decade, has found its way in the field of computational motion analysis. A kinematic based Expressive MOTion Engine (EMOTE) has been developed by Norman Badler's Group [10]. Swaminathan et al. [11] propose a probabilistic model which uses a body kinematic model and joint velocities to model Shape qualities. Santos and Dias address trajectory-based Laban models [12], while Kamrad and Dias focus on body part acceleration signals for Laban-based behaviour understanding [13]. Kim et al use an RGB-D camera to extract joint velocities to model the Effort component [14]. Zhao [15] and Rett [16] both investigated LMA in the context of communicative gestures. Zhao explored inverse kinematics, while Rett exploited vectorial information of limb velocity and acceleration signals. The use of classical visual cues has not yet been addressed, from which an unsolved challenge is identified and pointed as a valuable contribution.

1.2 Our Approach

This manuscript is presented as a contribution to computational LMA and Person Recognition areas. In this latter area, most existent applications are based on biometric properties such as fingerprint, eye scanning or face recognition technologies. Non-invasive, motion-based biometric systems are still very dependent on a specific activity, gait. This work is an extended solution to the approach proposed by Santos and Dias [2]. We address two unsolved problems in the current state of the art: (1) Existent LMA models are mostly based on kinematic or motion dynamic features, whereas classic visual cues have not yet been applied to this purpose; (2) Despite the advances in motion based person recognition frameworks in recent years, gait is still the dominant exploited activity. We illustrate our approach in Figure 1. We propose to acquire a set of image sequences from a calibrated camera network, from which silhouettes are segmented. Given their high dimensionality, we propose an alternative representation, which combines two signal processing techniques: Principal Component Analysis and Fast Fourier

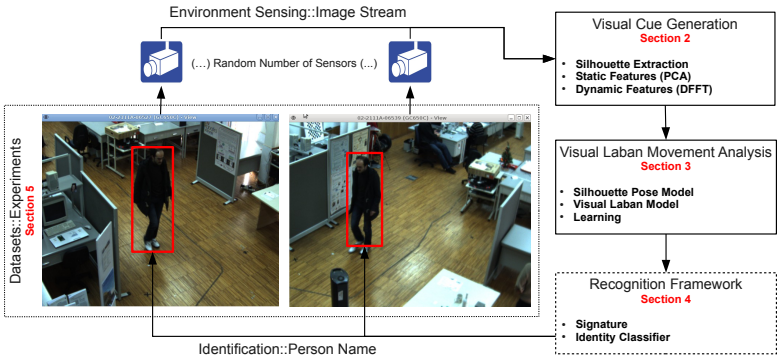


Fig. 1: Simplified block diagram of the proposed framework.

Transform. Posteriorly, we apply a supervised learning strategy to associate the generated silhouette features to sets of training data, which are manually labelled with the dominant LMA characteristics. Upon learning the Visual LMA models, we use a classifier based on Dynamic Bayesian Networks, which is applied to autonomously analyse motion sequences using LMA symbolic descriptors. This motion analysis framework is integrated with an adapted version of the person identification system developed in [2]. By solving the aforementioned challenge, we identify the following major contributions: (1) A LMA model based on visual cues; (2) Establish an adapted Laban signature model for action-based person identification. Validation is done on two publicly available datasets, through adequate evaluation metrics. Additional experimental information will be found on a support web page[3].

1.3 Paper Structure

In Section 2 we address silhouette encoding with the purpose of generating features representing static and dynamic properties of motion, from which the relevant results and considerations are presented. The learning and inference strategies for the LMA model are explained and tested in Section 3. We integrate our developed Vision-based LMA models in the person recognition framework in Section 4, introducing the signature and identity model strategies and results. Conclusions are discussed in Section 5.

2 Visual Cues and Variables

In this work, we are encoding classical visual features, which are then used to learn motion models. For the purpose of extracting silhouettes in our experimental set-up, we have applied existing image processing algorithms, from which we

[3] `http://www.isr.uc.pt/~luis`

mainly enumerate the popular Gaussian filtering, background subtraction and Canny Edge Detector.

2.1 Silhouette Features

Consider a binary image I_{BW}, containing a silhouette contour P, which is represented by the coordinates of every white pixel, such that:

$$P = \begin{bmatrix} (u_1, v_1) \\ \vdots \\ (u_s, v_s) \end{bmatrix}, \forall\ (u_s, v_s) \in I_{BW} : I_{BW}(u_s, v_s) = 1 \tag{1}$$

Given the potential high dimensionality of P, we use an alternative representation based on Principal Component Analysis (PCA) algorithm. PCA is used to uniquely characterize the internal geometrical structure of the scatter data, which best explains its variance. The defined orthogonal PCA space is here hypothesized as a set of **geometrical** cues. In practice, we use the eigenvector coordinates $v_r = (x_r, y_r)$, (which define the axis in the PCA component space) as independent silhouette features p for a static image I^t_{BW} at instant t. We compute the ratio between the first and second eigenvalues as λ_1/λ_2, which is then multiplied by the first component coordinates, such that $v'_1 = v_1(\lambda_1/\lambda_2)$. This additional step makes the first eigenvector to represent implicit eigenvalue information, which mitigates the impact of silhouette scale. The second component coordinates are used directly.

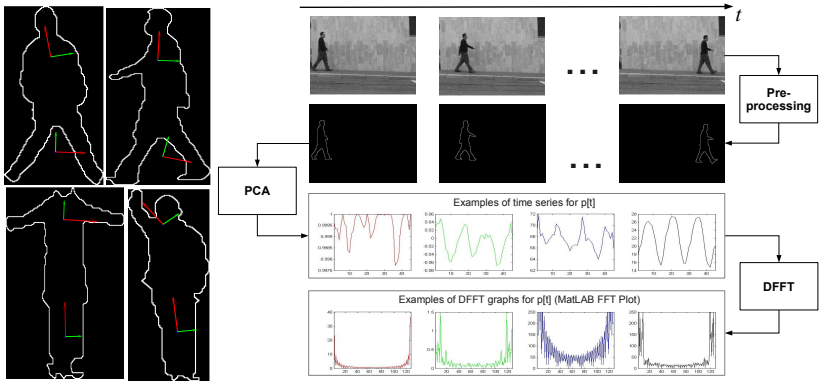
$$\hat{P}^t = pca(P) : \hat{P} = [v'_1\ v_2] \equiv [p^t_1, \cdots, p^t_4] \tag{2}$$

This procedure is applied twice, in both upper and lower body sections of the silhouette, which is roughly divided using the information about its center of mass. We consider this approach to provide better information in cases where actions are dominantly performed by one of the body halves (e.g. run or wave).

A part of LMA components address movement expressiveness, which describe **motion dynamics**. These properties require temporal characterization, rather than using a single image. Let an image sequence $\mathbf{I}$ be divided into sub-sequences of duration $\hat{t}$, such that $\hat{\mathbf{I}} = \{I_{t-\hat{t}}, \cdots, I_t\}$. For each $\hat{I}$ we have a corresponding times series for each feature p_j, such that $p_j[\hat{t}] = (p^{t-\hat{t}}_j, \cdots, p^t_j)$. To characterize motion dynamics we apply the Fourier Transform, a popular technique to analyse time series. We propose represent $p_j[\hat{t}]$ in the frequency domain (eq. (3)), from which the computed coefficients constitute a set of features, implicitly representing the dynamics of $\hat{\mathbf{I}}$.

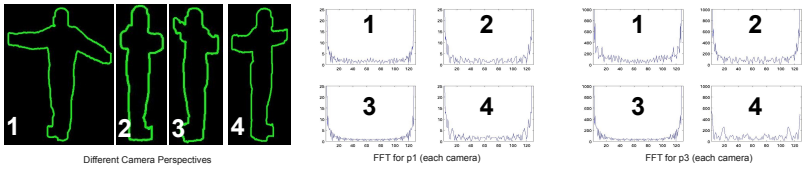
$$P_j(\omega) = \sum_n p_j[\hat{t}] e^{-i\omega n} \tag{3}$$

Within the set of Fourier coefficients, we select the maximum value, such that $F = \{f_1, \cdots, f_j\}\ :\ f_j \in \max P_j(\omega)$, and its correspondent fundamental frequency index n. The last considered feature is based on the **silhouette displace-**



(a) Examples of Static features.

(b) Processing diagram of a sequence $\hat{\mathbf{I}}$.



(c) Dynamic features for $p_1(\hat{t})$ and $p_3(\hat{t})$, in 4 different perspectives.

Fig. 2: Static (a), Dynamic (b) visual cues examples and an example of Dynamic for 4 different perspectives.

ment vector $\overrightarrow{d}$ for two consecutive images, such that $\overrightarrow{d}_t = (u_c^t - u_c^{t-1}, v_c^t - v_c^{t-1})$, where (u_c^t, v_c^t) is the center of mass at instant t. The displacement feature is $\theta_t = atan2(v_c^t - v_c^{t-1}, u_c^t - u_c^{t-1})$, which will be specially relevant for components which are direction based, rather than orientation, being complementary to the information given by the PCA features.

2.2 Experimental Results

In this section, for simplicity and easier visualization, results are presented for a representative subset of features $\{x'_1, y'_1, {}^{\lambda_1}/_{\lambda_1}, \} \in \hat{P}, \theta$ and $\{f_1, f_2\} \in F$, for the different gestures g_{index} in the selected datasets. Figure 3 presents the mean of measured values for each feature in each different gesture. Results show features to exhibit discriminant capabilities with respect to different gestures and, consequently, different properties of the performed motion. These differences indicate that features encompass interesting geometric and dynamic properties to be explored by our model, in Section 3. We also highlight an interesting fact observed during a preliminary study on action observation from different perspectives. The bottom image in Figure 2 shows that silhouettes are naturally different, however their feature dynamic behaviour is similar in the various perspectives.

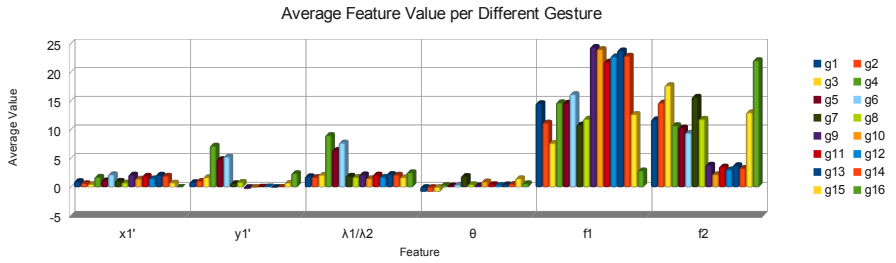


Fig. 3: Average values for features $\{v_1', \lambda_1/\lambda_2, \theta, f_1, f_2\}$, computed across 16 gestures g, belonging to the datasets used in our work.

3 Visual Laban Model

The computed feature variables are used to define the LMA model, which is learned using a supervised mixture model approach. Each component will be modelled from adequate feature types. A Bayesian classifier is used to evaluate the model with respect to its analysis capabilities, which is required to accurately characterize short activity sequences.

Definitions: The proposed LMA model is a probabilistic representation of its components, learned from a set of observable visual-based cues $\{f_i,\theta,p_i\}$, which are extracted from a combination of static images and sequences. It is parametrised into as many sub-models as the number of components c_n, where the initial hypothesis for the space state is defined in equation (4) based on the concepts of Labanotation [17] (details in subsection 3.1). In light of the conclusions withdrawn from the feature experiments, the model should be prepared to support an arbitrary number of cameras S_m, therefore bear in mind the pro-

Program
- Description
  - Specification
    - **Relevant Variables:**
    - $c_n \in \mathcal{L}$: represents the n^{th} Laban component;
    - $p_j \in \hat{P}$: represents the j^{th} static silhouette feature;
    - $f_i \in F$: the i^{th} dynamic silhouette feature;
    - $\theta \in [-\pi, \pi]$: represents the relative displacement of the silhouette;
    - **Decomposition:**
    - $P(c_n, p_j, f_i, \theta) = P(c_n)P(p_j|c_n)P(f_i|c_n)P(\theta|c_n)$
    - **Parametric forms:**
    - $P(c_n) = \begin{cases} \text{uniform} & , t = 0 \\ P_{t-1}(c_n) & , t > 0 \end{cases}$
    - $P(p_j, f_i, \theta|c_n)$: independent Gaussian distributions;
  - **Identification:** none;
- **Questions:**
- $P(c_n|f_i, p_j, \theta) : P(c_n|f_i), P(c_n|p_j), P(c_n|\theta)$

Fig. 4: Bayesian Program for estimating the most probable state for each Laban component c_n. It considers a single component, for simplicity purposes.

posed model is assumed to be running independently for each different sensor. In practice, each camera will show a different silhouette perspective.

$$\mathcal{L} = \begin{cases} c_1 \text{ :Effort Time} & \in \{sudden, sustained\} \\ c_2 \text{ :Effort Space} & \in \{direct, indirect\} \\ c_3 \text{ :Effort Flow} & \in \{free, bound\} \\ c_4 \text{ :Shape Form} & \in \{wall/pin, ball\} \\ c_5 \text{ :Direction Shape} & \in \{spoke, arc\} \\ c_6 \text{ :Shape X} & \in \{spreading, enclosing\} \\ c_7 \text{ :Shape Y} & \in \{rising, sinking\} \\ c_8 \text{ :Shape Z} & \in \{advancing, retreating\} \end{cases} \tag{4}$$

Formulation: The Bayesian Program in Figure 4 shows our proposed *visual* LMA model. The first step is to state and define the relevant variables:

- $\mathcal{L} = \{c_n \equiv \{q_1, q_2\}\}$ is a variable denoting a LMA component representing a specific motion characteristic, admitting two mutually exclusive states, as defined in equation (4). States are assumed to be dynamically propagating through a given sequence.
- $F = \{f_i \in \mathbb{R}_0^+\}$ are a set of random variables representing dynamic information from an image sequence $\hat{I}$.
- $\hat{P} = \{p_j \equiv \{u_1, \cdots, u_x\}\}$ are a set of random variables representing the geometrical information over static silhouettes $\in I$. It is discretised into a number of x equidistant intervals, representing a possible state (or bin) u_x.
- $\theta \in [-\pi, \pi]$ is a random variable which represents the displacement orientation between to consecutive silhouettes P^t and P^{t-1}.

The model decomposition is detailed in the Bayesian Program in Figure 4, from which the questions are formulated. To estimate of a given state in the model, Bayesian inference is applied. Inference in this work considers feature variables to be independent and identically distributed, and can be formulated, from a *Maximum a Posteriori* perspective, as follows:

$$\begin{aligned} P(c_n|p_j, f_i, \theta) &\propto P(c_n) \textstyle\prod_{\forall i,j} P(f_i, p_j, \theta|[c_n = q_j]) \\ &\propto P(c_n) P(\theta|c_n) \textstyle\prod_i P(f_i|c_n) \prod_j P(p_j|c_n) \end{aligned} \tag{5}$$

3.1 Laban Component Models and Feature Types

LMA is a symbolic notation language used for a comprehensive understanding of human motion, which has the unique capability to describe expressiveness. It was developed around the concepts of movement notation and integrates studies from anatomy, kinesiology, psychology and Labanotation. It was originally designed for dance choreography, and is today one of the most used systems for movement analysis in a wide range of areas. It defines movement as an intentional process of patterned and orderly changes, which are better studied if divided in different levels. It is divided in four main different components, each describing a different motion property, using an adequate symbolic grammar, Labanotation [17]. *Body* and *Space* components describe the structural or physical properties of the body and spacial patterns along with body part pathways respectively. The *Effort*

component addresses dynamics and inner intention, while the *Shape* deals with the connections between the body and space and the changes in body shape. We hypothesize a generalization of LMA to full body analysis rather than specific body parts, which may, for some activities, present undefined states (e.g. rising symbolizes a state which may not even be observable for some activities). In this work, we model the two components addressing motion expressiveness, *Effort* and *Shape*, which are always observable in any movement sequence [15].

Effort addresses the dynamic properties of motion with respect to inner intention. It is divided into four different qualities: *Space*, *Time*, *Weight* and *Flow*. We discard *Effort Weight* as it is usually associated to strength and we consider that visual cues are not adequate for its characterization.

- c_1 *Effort Time* characterizes the cognitive process of decision, which is tightly related to time. Therefore it is associated to dynamic features f_i and the Bayesian question yields $P(c_1|f_i)$.
- c_2 *Effort Space* is focused on the attention with respect to *orientation with a purpose*. We hypothesize this to be a combination of geometric, dynamic and orientation features, such that $P(c_2|f_i, p_j, \theta)$.
- c_3 *Effort Flow* characterizes motion continuity, which is related to performance along time and whether or not it is contained to a single action. Dynamic and displacement are the selected feature types, hence $P(c_3|f_i, \theta)$.

Shape emerges from the *Body* and *Space* components. As the name implies it address the geometrical form the body takes and how it changes in time. It is used to integrate different categories into movement.

- c_4 *Shape Form* is one of the categories of Shape, is as the name implies is the form the body takes, which is mostly geometric, such that $P(c_4|p_j)$. We simplified the space state considering wall and pin as a single state, where the performer is dominantly standing.
- c_5 *Directional Shape* represents the existent relation between the body and the environment. It divides movements into spoke-like (e.g. point) and arc-like (e.g. waving bye bye). It is mostly geometric and $P(c_5|p_j)$.
- $c_{6,7,8}$ Shape also has qualities, which describe body extensions or how its form changes with respect to specific spacial orientations. Geometric features are relevant, as is the way these change in time, with respect to the body center. Thus, the Bayesian question formulates as $P(c_6, c_7, c_8|f_i, p_j, \theta)$.

3.2 Learning the Models from Experimental Data

The likelihood distributions $P(\theta, p_j, f_i|c_n)$ represent the actual LMA model based on sets of training motion sequences. The first step in the learning process is the manual annotation of the image sequences that will be used to train the model. Let a sequence $\mathbf{I}$ be labelled with a set of dominant LMA states, one for each corresponding component c_n. For each $\hat{\mathbf{I}} \in \mathbf{I}$ we compute a set of features $\{p_j, f_i, \theta\}$. For every component c_n, we cluster the different features with

respect to each possible state, i.e. two different sets of features are associated to classes q_1 and q_2. Hence, considering features as independent and identically distributed, for a given state q_k we have a likelihood distribution defined from the association process as:

$$P(\beta|[c_n = q_k]) = \mathcal{N}(\mu_\beta, \sigma_\beta),\ \beta \in \{f_i, \theta\} \tag{6}$$

In the specific case of p_j we have discretised it in a number x equidistant intervals, between observed range for p_j. Each interval corresponds to a single possible state u, such that, $p_j = \{u_1, \cdots, u_x\}$. Hence, the likelihoods using this variable formulate as a stochastic matrix $M^{c_n}_{x,r}$, each cell's probability is given by:

$$P([p_j = u_x]|c_n = q_k) = \frac{\sum \text{observations } u_x \text{ for state } q_k}{\sum \text{total observations } u \text{ for state } q_k} \tag{7}$$

3.3 Experiments on Motion Characterization with Laban Notation

The top two rows of Table 1 summarize the percentage of correctly identified states per component when compared to ground truth annotation and the average model confidence, i.e. the average probability with which states are classified. The average *f.a.r.* percentage is 87.67%, whereas the average confidence is over 90%. However, for component c_8, the accuracy is low with high confidence, which means the model is converging to a false positive for that particular characteristic. Isolating c_8 from the remainder, results show an accurate model when comparing ground truth annotation and estimated LMA states, showing that visual features can be applied for characterizing motion sequences using LMA

Table 1: Classified dominant LMA states. Each state is considered dominant, if and only if they are classified in at least 2/3 of the frames in a sequence *I*. The acronyms *f.a.r* and *a.m.c.* stand for Frame Accuracy Ratio and Average Model Confidence respectively. The acronym *n.d.* stands for Non-Dominant state.

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
f.a.r.%	94.02	83.17	91.43	67.02	80.34	81.91	63.17	21.43
a.m.c.%	87.42	93.18	90.58	82.88	89.24	98.91	90.46	90.88
bend	sustained	indirect	free	arc	n.d.	n.d	sinking	n.d.
jack	n.d.	n.d.	free	n.d.	n.d.	n.d.	rising	n.d.
jump	sudden	direct	free	spoke	pin	n.d.	rising	advancing
pjump	sudden	direct	free	spoke	pin	n.d.	rising	n.d.
run	sudden	direct	free	spoke	pin	enclosing	rising	n.d.
side	sudden	direct	free	spoke	pin	spreading	n.d.	n.d.
skip	sudden	direct	free	spoke	pin	n.d.	rising	n.d.
walk	sustained	direct	free	spoke	pin	n.d.	rising	advancing
wave1	sustained	indirect	free	arc	wall	spreading	n.d.	n.d.
wave2	sustained	indirect	free	arc	wall	spreading	n.d.	n.d.
boxing	sudden	direct	bound	spoke	nd	spreading	n.d.	advancing
handclapping	sustained	indirect	bound	bound	n.d.	n.d.	n.d.	n.d.
handwaving	sustained	indirect	free	n.d.	ball	spreading	rising	n.d.
jogging	n.d.	n.d.	free	spoke	n.d.	n.d.	n.d.	n.d.
running	n.d.	n.d.	free	spoke	n.d.	n.d.	n.d.	advancing
walking	n.d.	n.d.	free	spoke	pin	enclosing	n.d.	n.d.

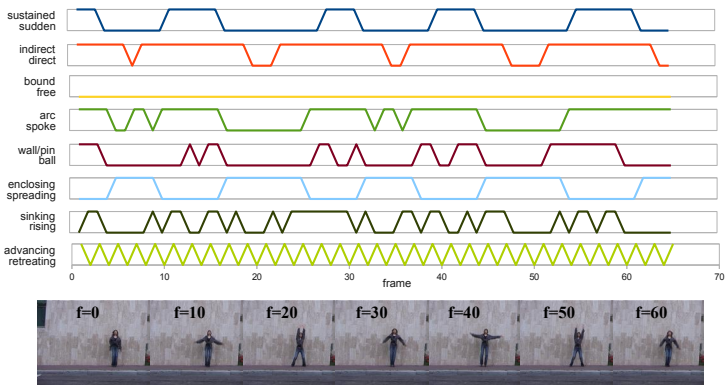


Fig. 5: Example of Laban symbolic classification ($q_1 \oplus q_2$) for gesture "jumping jack", performed by "daria" from Weizmann dataset and some key frames.

descriptors. To further extend our result analysis, Table 1 additionally presents the dominant characteristics for each of the analysed activities. Gestures are usually characterized by specific properties, which using our model, are accordingly represented by dominant symbols, e.g. actions traditionally associated to fast movement are associated to the symbol *sudden*. Figure 5 presents an example of the generated symbolic output of the visual-based LMA model.

4 Application on Person Recognition

Santos and Dias [18] state that LMA can be generalized, at least symbolically. In their research, as in this work, LMA states are shown to be repeatable for similar actions when performed by different persons. However, the confidence with which the model classifies each state, differed from person to person. This property indicates that LMA space can be discriminant with respect to whom is performing an observable sequence. The presented identification model is different from the one presented [18], and has been adapted to cope with the generalization characteristics of the proposed visual LMA model.

4.1 Laban-based Motion Signature and Identification Models

Laban Space Definition [2]: Let LMA Space $\chi \in \mathbb{R}^n$ be a n-dimensional unified representation for all LMA variables c_n. Consider a vector $R \in \chi : R = (\tau_1, \cdots, \tau_n)$ representing a combination of LMA properties for a motion subsequence, where $\tau_n = P(c_n = q_1)$. Consider each LMA variable as a topological node, measuring inter-node distances as $d_{i,j} = \tau_i - \tau_j$. The signature is generated by defining the adjacency matrix A of the topological graph and computing its single value decomposition. The signature variables $\Gamma = \{\gamma_i\}$ are independent and defined from the computed eigenvalues and corresponding eigenvectors.

Signature and Person Identification Model: In [2], Laban is applied to analyse different body parts rather than the body as a whole, which naturally augments the signature discriminant capabilities. The initial experiments applying our visual LMA model to the identification model in [2], presented poor recognition accuracy. Result evaluation showed that, generalizing LMA to a full body, lead to similar estimates, in Laban space, for the different persons in the same action categories, which propagated to the identification model, causing a high rate of misclassified identities. To overcome this issue, we propose a modified version of their identification model, adding an extra variable to the signature model, activity $\alpha \in \Lambda$, with the purpose of increasing discrimination. Let each $\gamma_i \in \Gamma$ be an independently and identically distributed motion signature feature. Consider the following joint distribution for the identification model:

$$P(\zeta, \alpha, \gamma_i, c_n) \tag{8}$$

where ζ represents a recognition variable, whose states correspond to different identities. We consider the following decomposition:

$$P(\alpha|c_n)P(c_n, \alpha|\gamma_i)P(\alpha, \gamma_i, c_n|\zeta)P(\zeta) \tag{9}$$

The prior distribution $P(\zeta)$ starts as an uniform distribution in the first iteration. The activity Λ is estimated combining the different LMA variables. In our adapted approach, we learn a signature kernel model $P(\gamma_i|c_n, \alpha)$, which is now indexed to specific activities. This means that, for each activity $\alpha \in \Lambda$, a signature Γ is computed for each different person, creating a kernel upon computation for all $\alpha \in \Lambda$. Using Bayesian inference, the posterior density yields:

$$P(\zeta|\alpha, \gamma_i, c_n) \propto P(\zeta) \prod_{q=1}^{n} P(c_q|\alpha) \prod_{q=1}^{n} P(\alpha, c_q|\gamma_i) \prod_{q=1}^{n} \prod_{p=1}^{i} P(c_n, \gamma_p, \alpha|\zeta) \tag{10}$$

The normalization factor is omitted for simplification purposes. The distributions $P(\alpha, \gamma_i, c_n|\zeta)$, $P(c_n|\alpha)$ and $P(\alpha, c_n|\gamma_i)$ are the likelihood distributions, representing the identity model trained from real experimental data. The distributions $P(c_n|\alpha)$ are Gaussian, while $P(\alpha, c_n|\gamma_i)$ is a kernel of Gaussian distributions for γ_i generated from the probability values of c_n and indexed by α. The identity likelihood is a multi-variate stochastic matrix where signatures are associated to identities by means of activity and LMA state indexing.

4.2 Experimental Set-up and Results

Our identification experimental set-up encompasses 2 acknowledgeable datasets in motion analysis, KTH[4] and WZ[5]. **As far as the author's knowledge goes, this is the first study in which these datasets are used for Person Recognition.** A LMA description is generated for each motion sequence,

[4] [Online] `http://www.nada.kth.se/cvap/actions/`
[5] [Online] `http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html`

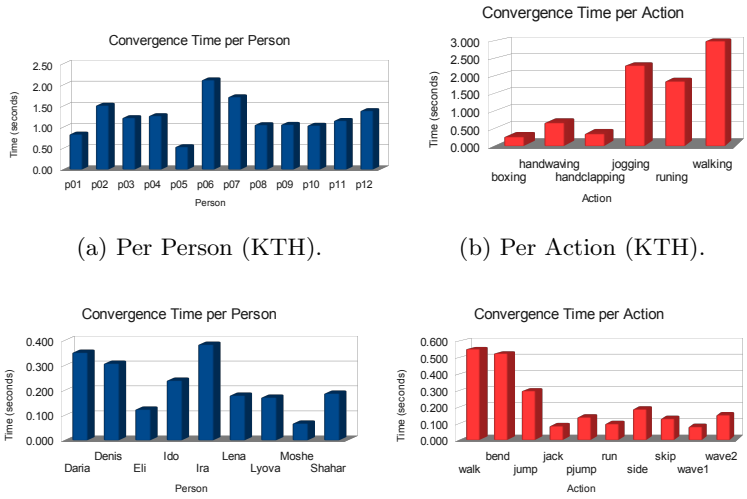
	p01	p02	p03	p04	p05	p06	p07	p08	p09	p10	p11	p12	per-seq.(%)
person01	0.85	0.01	0	0.03	0	0.07	0	0.02	0	0	0.02	0	100.00
person02	0	0.92	0.03	0	0	0	0	0	0.01	0	0.03	0	100.00
person03	0	0	0.95	0.02	0.01	0	0	0	0.02	0	0	0	100.00
person04	0	0	0.01	0.93	0	0	0	0	0.04	0	0.01	0.01	100.00
person05	0	0	0	0.02	0.97	0	0	0.01	0	0	0	0	100.00
person06	0	0	0.02	0.01	0	0.92	0.02	0	0	0	0	0	100.00
person07	0	0	0.01	0.02	0	0.03	0.94	0	0	0	0	0	100.00
person08	0	0	0	0.01	0.01	0	0	0.96	0.01	0	0	0	100.00
person09	0	0	0	0.05	0	0	0	0	0.94	0	0	0	100.00
person10	0	0	0	0.01	0.02	0	0	0.01	0.01	0.94	0	0.02	100.00
person11	0	0.01	0.02	0.01	0	0	0	0	0.01	0	0.94	0	100.00
person12	0	0	0	0.03	0	0	0.01	0.01	0.02	0	0	0.93	100.00

Fig. 6: Results for identification on KTH dataset: Confusion matrix for per-frame classification accuracy; the last column shows per-sequence classification accuracy. (overall accuracy = 100%).

which is used to generate signatures and identity classification. We are interested in knowing the identification rate, i.e. how many times is a person correctly identified. Classification results are presented on a per frame and per sequence basis. The **KTH dataset** has 6 different actions (walking, jogging, running, boxing, hand waving, hand clapping)acquired at 25 *fps* frame rate and a 4 second average length. We have used 12 different actors. Identification results for the KTH dataset show an average per-frame accuracy over 90% and a perfect per sequence accuracy (Fig.6). Convergence is typically reached in 1 and 2 seconds time, considering the last instant where a misclassified frame was observed. In fact, analysing per action results, we see the major confusion focus lies in similar actions, for which the classifier takes longer to converge. The **WZ dataset** is a collection of 90 video sequences of low resolution 180×144), recorded at 50 *fps* frame rate. It has 9 different people, performing 10 different actions (run, walk, skip, jumping jack, jumping, jump in the place, side, waving with one and two hands, bend). As in the previous dataset, per-sequence classification achieved 100% accuracy (Fig.7). The per-frame ratio is still high, which is positive due to the number of similar actions. The convergence times are faster than in KTH, which is justified by an increased *fps* acquisition rate.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	per-sequence(%)
(1) Daria	0.82	0.02	0	0.06	0.01	0.01	0.06	0.02	0.01	100.00
(2) Denis	0.08	0.78	0	0.05	0	0.08	0.01	0.01	0	100.00
(3) Eli	0.02	0	0.92	0.01	0.01	0	0.02	0.01	0	100.00
(4) Ido	0.10	0	0	0.80	0.01	0.02	0.02	0.04	0	100.00
(5) Ira	0	0.01	0.02	0.02	0.84	0.03	0.02	0.06	0.01	100.00
(6) Lena	0.05	0.01	0	0	0.01	0.91	0.01	0	0	100.00
(7) Lyova	0.01	0	0	0.03	0	0.08	0.86	0.02	0	100.00
(8) Moshe	0	0	0.01	0	0.01	0.03	0	0.95	0	100.00
(9) Shahar	0.02	0	0.01	0.01	0.01	0	0.06	0.01	0.87	100.00

Fig. 7: Results for identification on Weizmann dataset: Confusion matrix for per-frame classification accuracy; the last column shows per-sequence classification accuracy. (overall accuracy = 100%).



(a) Per Person (KTH). (b) Per Action (KTH).

(c) Per Person (WZ). (d) Per Action (WZ).

Fig. 8: Average convergence time (seconds), considering a video sampled with fps = 25 and fps = 50 frames per second for KTH and WZ datasets respectively.

5 Conclusions and Future Work

In this work we have developed a model for a comprehensive characterization of motion sequences, which uses visual cues and is supported by an adequate descriptive grammar, based on the principles of LMA and notation. The analysis framework was integrated in a previously developed person recognition framework, for which the models were adapted to cope with the LMA generalization to the body as a whole. The generated LMA symbolic description was applied into activity identification and to develop motion signatures, which are combined into showing discriminating capability between different persons, using a Bayesian classifier. Results are promising and suggest further exploitation and model development. We intent to explore the feature behaviours with respect to different acquisition perspectives, by augmenting the model to multiple cameras simultaneously. It is our expectation to improve the signature model into relaxing the specific activity dependency. Furthermore, we aim to continue validating our identity estimation accuracy in more complex datasets, improving image segmentation and provide a working prototype.

Acknowledgements. Luís Santos is supported by FCT - Portuguese Foundation for Science and Technology, Grants # 65935/2009. This work has been supported by Institute of Systems and Robotics from University of Coimbra, Portugal and Khalifa University, Abu Dhabi, UAE.

References

1. Cornelius Held, Julia Krumm, Petra Markel, and Ralf Schenke. Intelligent video surveillance. *Computer*, March:83–84, 2012.
2. Luís Santos and Jorge Dias. Person identification based on activity invariant signatures. (under peer review) on Journal of Pattern Recognition Society, Elsevier.
3. Irmgard Bartenieff and Dori Lewis. *Body Movement: Coping with the Environment.* Gordon and Breach Science, New York, 1980.
4. Z Liu, L Malave, and S Sarkar. Studies on silhouette quality and gait recognition. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on (Volume:2)*, 2004.
5. X. Hou and Z. Liu. Fusion of face and gait for human recognition in video sequences. In *Int. Conf. on Information Technology and Computer Science*, 2009.
6. S. Lee, Y.i Liu, and R. Collins. Shape variation-based frieze pattern for robust gait recognition. In *IEEE Conf. on Comp. Vision and Pattern Recognition*, 2007.
7. C BenAbdelkader, R Cutler, and L Davis. Stride and cadence as a biometric in automatic person identification and verification. In *Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, 2002.
8. T Kobayashi and N Otsu. Action and simultaneous multiple-person identification using cubic higher-order local auto-correlation. In *Proceedings of the 17th International Conference on Pattern Recognition*, 2004.
9. Alexandros Iosifidis, Anastasios Tefas, and Ioannis Pitas. Activity-based person identification using fuzzy representation and discriminant learning. *IEEE Trans. on Information Forensics and Security*, 7(2):530–542, 2012.
10. Diane Chi, Monica Costa, Liwei Zhao, and Norman Badler. The emote model for effort and shape. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '00, pages 173–182, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.
11. Dilip Swaminathan, Harvey Thornburg, Jessica Mumford, Stjepan Rajko, Jodi James, Todd Ingalls, Ellen Campana, Gang Qian, Pavithra Sampath, and Bo Peng. A dynamic bayesian approach to computational laban shape quality analysis. *Adv. in Hum.-Comp. Int.*, 2009:2:1–2:17, January 2009.
12. Luis Santos and Jorge Dias. Motion patterns: Signal interpretation towards the laban movement analysis semantics. In *2nd Doctoral Conference on Computing, Electrical and Industrial Systems*, 2011.
13. Kamrad Khoshhal and Jorge Dias. Probabilistic human interaction understanding - exploring relationship between human body motion and the environmental context. *Pattern Recognition Letters*, 34:820–830, 2013.
14. Woo Hyun Kim, Jeong Woo Park, Won Hyong Lee, Hui Sung Lee, and Myung Jin Chung. Lma based emotional motion representation using rgb-d camera. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, HRI '13, pages 163–164, Piscataway, NJ, USA, 2013. IEEE Press.
15. Liwei Zhao. *Synthesis and Acquisition of Laban Movement Analysis Qualitative Parameters for Communicative Gestures.* PhD thesis, Univ of Pennsylvania, 2002.
16. Joerg Rett. *Robot - Human interface using laban movement analysis inside a bayesian framework.* PhD thesis, University of Coimbra, 2009.
17. Ann H. Guest. *Labanotation or Kinetography Laban.* Theatre Arts, N.Y., 1970.
18. Luís Santos and Jorge Dias. Laban-based multilayer model for activity recognition and annotation. (under peer review) on International Journal on Information Sciences, Elsevier.