



## Versão Final do Projecto

( <http://www.vmnf.net/vmnf/ist/meic/cadeiras/12-immi/projecto/> )

### Navegador Multimodal de Imagens (NavMI)

Vitor M. N. Fernandes

vmnf@yahoo.com

IST/DEI: MEIC: IMMI

(v3.00 - 2005/07/11)



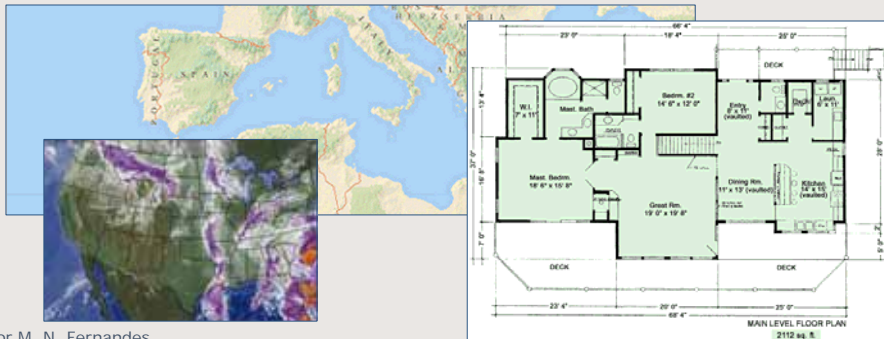
## Nav MI Agenda

- **Projecto NavMI**
  - O que é?
  - As tarefas possíveis
  - Resultados da avaliação do Protótipo Funcional 2
- **Versão Final do Projecto**
  - Funcionalidades
  - A modalidade voz
  - Redesenho da Interface e Ajuda disponível
  - **Demonstração**
  - Análise de Tarefas e Problemas detectados,
  - Conclusões
  - Trabalho futuro



## Nav MI O que é?

- É um aplicação que usa três modalidades de comando (teclado, ponteiro e voz) para **visualização e navegação** em imagens de alta resolução 2D como sejam:
  - Fotos de satélite, informação geográfica, plantas de prédios,...



Vitor M. N. Fernandes  
IST/DEI: MEIC: IMMI

3  
Julho 2005



## Nav MI As tarefas possíveis...

- Os comandos podem ser inseridos / complementados através de 3 modalidades:
  - **Digitar** (teclado), **apontar** (rato) e **voz** (fala).
- Ao nível das possibilidades de comando destaco:
  - Comandos de navegação relativa na imagem:
    - p.e.: esquerda, direita, baixo e cima e escala (zoom in / out); ✓
  - Comandos de navegação absolutos:
    - p.e.: esquerda tudo ou um determinado nível de zoom (zoom 60%); ✓
  - Comando de marcação de locais: ✗
  - Comandos para locais como centrar num local: ✗
  - Comandos de animação: ✗
  - Reconhecimento de voz ✓
    - Modo 1: comando voz isolado (pressionar botão reconhecer)
    - Modo 2: comandos voz em contínuo
      - Submodo 2.1: Obrigatório "Executar comando" <comando>
      - Submodo 2.2: Após "Activar modo comando" basta <comando>

Vitor M. N. Fernandes  
IST/DEI: MEIC: IMMI

4  
Julho 2005



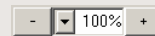
## Nav MI Avaliação do PF2

- 4 Utilizadores a emitir um comando de cada tipo:
  - Direita, cima, baixo, esquerda
  - Tudo direita, tudo cima, tudo baixo, tudo esquerda
  - Zoom in, zoom out, zoom 10%, zoom 50% e zoom 100%
- Resultados
  - Detectado um bug na gramática
    - Quando "zoom <valor>" apenas considerava zoom desprezando outros sinónimos
  - Detectadas novas palavras a incluir na gramática:
    - Virar, percorrer, caminhar, chegar, distanciar
    - **!!! Mas que quando utilizadas geraram o mesmo tipo de comando!!!**
  - Apenas 25% de erros na detecção de **oeste -> leste?!...**
    - Foram experimentadas várias soluções mas optou-se pelos sons
      - éste | léste | ôéste
  - Média de confiança nos comandos emitidos reconhecidos correctamente
    - **75%de confiança (sempre > 60%)!**




## Nav MI Versão final do projecto

- Funcionalidades:
  - **Viewfinder:**
    - Evolução...** ■ Aceita limite em toda a margem (imagem a visualizar < área disponível)
  - **Implementação do Zoom:**
    - Evolução...** ■ Imagem usada para 100% a resolução 3795x2645
    - Pré-processamento das 10 imagens (10 zooms)
      - Bicubic resize de imagem 7590x5290!
    - 10 valores de escala: 100%, 90, 80%,... 20% e 10%
      - Aceita tamanhos menores do que a área total disponível
    - Limites colocados nas imagens
      - Margem a negro para o utilizador ter a pista visual de quando se encontra nos limites da imagem em que navega
  - **Reconhecedor:**
    - Novo!** ■ O reconhecedor funciona em 2 modos




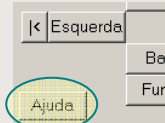


# Nav MI Versão final do projecto

- Modalidade voz: 
  - Estado do sistema visível** (Metáfora do semáforo)
    - Noval** Parado / a não reconhecer –Vermelho
      - Comandos directos (Submodo 1) – Verde
      - Comandos “**Executar comando**” <comando> (Submodo 1) - Amarelo
    - Modo 1:** Comando voz isolado  Modo contínua
    - Noval** É possível emitir um comando isolado e o reconhecedor pára de imediato
    - Modo 2:** Comandos voz em contínuo  Modo contínua
  - Evolução...** Com dois submodos...
    - Submodo 2.1:** É necessário emitir “Executar comando” antes do comando a realizar – modo robusto para reduzir o número de falsos positivos e para ser usado em condições com “muito ruído”
      - Entra-se neste modo dizendo “Activar modo comando”
    - Submodo 2.2:** Modo já existente na versão PF2 em que se podem emitir comandos de seguida – mais suscetível de falsos positivos e a ser usado em condições “baixo ruído”
      - Sai-se deste modo dizendo “Desactivar modo comando”



# Nav MI Versão final do projecto

- Redesenho da interface
  - Nova colocação dos botões**
    - Noval** A utilização do botões de comando não respeitava nenhuma metáfora
      - Os botões tinham tamanhos diferentes para funções iguais
        - Por questões espaciais ainda se verifica nos comandos **tudo** esquerda e direita
      - Respeita a metáfora “Rosa dos ventos”
    - Maximização da área visível da imagem (em **18,7%**)
      - De **710x620** para **766x686**
  - Ajuda e documentação disponível
    - Evolução...** Para a gramática
 
    - Noval** Para a aplicação em geral
 



## Nav MI Versão final do projecto



## Nav MI Análise de Tarefas

- 2 utilizadores x 2 experiências (submodo 1 e submodo 2)
  - (13 comandos x 2 utilizadores x 2 submodos)
- Fiabilidade do submodo 1 (VFP) vs. submodo 2 (PF2)
  - Verificou-se que com a experiência nunca ocorreu um falso positivo!
    - Pediu-se ao utilizador que falasse o que entendesse no submodo 1
  - Verificou-se também que a pontuação de confiança era maior no submodo 1 do que no submodo 2 (mais texto => maior confiança)
    - O utilizador tem que dizer antes de emitir um comando "Executar comando"
  - Submodo 1: **81,3%** (26) vs. Submodo 2: **75,0%** (26)
- "Activar modo comando" com uma taxa de **83%** (2) e
- "Desactivar modo comando" com uma taxa de **76%** (2)
- "Párar reconhecedor" com uma taxa de **68%** (4)
- Número de erros
  - considerando as expressões na gramática  $1/52=1,9\%$
  - geral  $3/52=5,8\%$



## Nav MI Análise de Tarefas

- 4 utilizadores x 1 experiência
- Tarefa 1 (simples): deslocar para a direita
  - Modalidade teclado
    - Tempo esperado: 5-10s, número de erros: 0
    - Tempo médio: 5,75s, número médio de erros: 0
  - Modalidade rato
    - Tempo esperado: 5-10s, número de erros: 0
    - Tempo médio: 7,5s, número médio de erros: 0,125
      - Confusão entre o tudo para a direita e o direita
  - Modalidade VOZ (modo 2 – submodo 2)
    - Tempo esperado: 5-20s, número de erros: 0
    - Tempo médio: 11,25s, número médio de erros: 0



## Nav MI Análise de Tarefas

- 4 utilizadores x 1 experiência
- Tarefa 2 (média): Ir para um canto da imagem
  - Modalidade teclado
    - Tempo esperado: 10-20s, número de erros: 0
    - Tempo médio: 19,5s, número médio de erros: 0
      - Com o teclado, por lapso, não é possível fazer "tudo \*\*!!!
  - Modalidade rato
    - Tempo esperado: 10-20s, número de erros: 0
    - Tempo médio: 12,25s, número médio de erros: 0
  - Modalidade VOZ (modo 2 – submodo 2)
    - Tempo esperado: 10-40s, número de erros: 1
    - Tempo médio: 17,5s, número médio de erros: 0,125



## Nav MI Análise de Tarefas

- 4 utilizadores x 1 experiência
- Tarefa 3 (complicada): mostrar Austrália com zoom a 80%
  - Modalidade teclado
    - Tempo esperado: 10-30s, número de erros: 2
    - Tempo médio: 18,5s, número médio de erros: 0,50
  - Modalidade rato
    - Tempo esperado: 10-30s, número de erros: 2
    - Tempo médio: 17,25s, número médio de erros: 0,25
  - Modalidade VOZ (modo 2 – submodo 2)
    - Tempo esperado: 20-60s, número de erros: 2
    - Tempo médio: 33,25s, número médio de erros: 1,0
      - Falar quando o reconhecimento termina e recomeça (50%)!
      - Mau reconhecimento (50%)!
  
- Voltar atrás ou passar por uma posição repetida foi considerado um erro!!!



## Nav MI Problemas detectados

- Problemas
  - Modalidade voz
    - Quase sem críticas ;-) !
    - Gramática
      - Pesquisar exaustivamente por sinónimos e inclui-los na gramática.
  - Modalidade apontar
    - Não suportar **zoom in / zoom out** com o rato (ficou em 2.º plano)
    - Não suportar o **arrastar** (várias problemas e eficiência?) (idem ^)
  - Modalidade teclado
    - Não suportar os comandos “tudo \*\*” – Bug detectado na versão final :-( !
  - Interface
    - Feedback voz pouco útil e que deve ser melhorado!



## Nav MI Conclusões

- Conclusões obtidas na realização do projecto
  - A modalidade voz:
    - é uma modalidade válida
    - é mais “lenta” do que apontar ou digitar
    - dá graus de liberdade que apontar e digitar não dão (largo o PC ;-) !)
    - “fugir” aos falsos positivos não é trivial e pode ser penalizante!
  - A geração de uma gramática válida é complicada e morosa
    - 2 questionários + entrevistas
  - Pouco tempo na passagem do PF2 para a VFP
    - Disco no sistema de desenvolvimento estragado > **24h**
    - Reinstalação do software da loquendo / reactivar licença >**24h**
    - Renovação da licença > **48h**
    - As duas semanas previstas passaram a uma :-(!
  - Os objectivos principais a que o projecto se propunha foram alcançados e validados por utilizadores “virgens” com 5 minutos (no máximo) de exploração / treino / explicação da interface.



## Nav MI Trabalho futuro

- Interface gráfica
  - Optimizar o funcionamento para imagens de grande dimensão!
  - Funcionamento de uma versão na PowerWall (OSGA?)
- Imagem
  - Receber apenas a maior imagem e uma descrição XML com os pontos dos vários locais / áreas
  - Pré-processar os vários níveis de zoom (p.e. bicubic resample)
  - Permitir a deslocação para pontos identificados na descrição XML
- Evoluções futuras
  - Geração dinâmica de gramática de modo a suportar novos locais de forma dinâmica
  - Gravar a nova descrição XML da imagem
  - Anotar na imagem... ;-) !



## Versão Final do Projecto

( <http://www.vmnf.net/vmnf/ist/meic/cadeiras/12-immi/projecto/> )

### **Navegador Multimodal de Imagens (NavMI)**

Vitor M. N. Fernandes

vmnf@yahoo.com

IST/DEI: MEIC: IMMI

(v3.00 - 2005/07/11)

