

CARACTERIZAÇÃO DE ALGORITMOS DE RASTREAMENTO DE OBJETOS EM VIDEO CONSIDERANDO SITUAÇÕES DE OCLUSÃO

Cleber de Souza Relli¹

Resumo

Este trabalho demonstra um estudo comparativo entre algoritmos de segmentação e rastreamento de objetos em vídeos, que é uma área de estudos da visão computacional. Foram utilizados três algoritmos de segmentação (Simples, Janelas Temporais e Métodos Estatísticos), combinados com quatro algoritmos de Rastreamento (por Predição, por Cores, Filtro de Kalman, e Fluxo Óptico). Foi implementado um *framework* o qual possibilita escolher o tipo de algoritmos de segmentação e o rastreamento utilizados. Para a realização dos testes foi utilizada a base PETS 2009 S2.L1 Time 12-34, que foi escolhida devido ao fato de conter cenas de objetos em *tracking*. No comparativo entre os algoritmos foi levado em consideração além do rastreamento dos objetos, a identificação dos que sofrem oclusão e qual o tipo que ocorre, definindo se há oclusão do tipo total, parcial ou mesmo não ocluso. Foram observados sessenta resultados, devido à combinação de três algoritmos de segmentação combinados com os quatro algoritmos de rastreamento sobre cinco vídeos existentes na base. Também é verificada a quantidade de objetos que entram e saem da cada cena, e a quantidade de pessoas pertencentes a cada frame. Além disso, avaliaram-se os resultados direcionados a ocorrência de oclusão, com intuito de explicar a causa dos erros na contagem dos frames que sofrem ou não oclusão. Os resultados obtidos pelo *framework* foram comparados com os dados da base rotulada manualmente. Como melhor resultado obteve-se com a combinação dos algoritmos de segmentação simples (subtração) com o algoritmo de rastreamento baseado no Filtro de Kalman, atingiu o menor índice de erros, sendo 38,20% para o conjunto de cinco vídeos.

Palavras-chave: 1. Segmentação, 2. Rastreamento, 3. Oclusão.

Abstract

This study demonstrates a comparative study of algorithms for segmentation and tracking objects in videos, which is an area of study in computer vision. We used three segmentation algorithms (Simple Temporal windows and Statistical Methods), combined with four tracking algorithms (for prediction, for Colors, Kalman Filter and Optical Flow). We implemented a framework which enables to choose the type of segmentation and tracking algorithms used. For the tests we used the 2009 base PETS 12-34 S2.L1 Time, which was chosen due to the facts scenes of objects tracking. In the comparison between the algorithms was taken into account beyond the tracking of objects, the identification of those who suffer from occlusion and which type is to define whether there is occlusion-type total, partial or no occlusion. Sixty results were observed due to the combination of three segmentation algorithms in combination with the four tracking algorithms about five videos in the base. It is also verified the number of objects in and out of each scene, and the amount of people belonging to each frame. In addition, evaluated the results targeted the occurrence of occlusion, aiming to explain the cause of errors in counting the frames suffering occlusion or not. The results were

¹ Professor da Faculdade Vizinhança Vale do Iguaçu - VIZIVALI; e-mail: cleber.relli@gmail.com.

compared with the framework with the data of the labeled manually. As best result was obtained with the combination of simple segmentation algorithms (subtraction) with the tracking algorithm based on Kalman filter, reached the lowest error rate, and 38.20% for all five videos.

Keywords: 1. Segmentation, 2. Tracking, 3. Occlusion.

1 - Introdução

A visão computacional é o conjunto de métodos e técnicas através dos quais sistemas computacionais podem ser capazes de interpretar imagens (WANGENHEIM; COMUNELLO, 2005).

Uma das possíveis formas de se utilizar visão computacional para aumentar o nível de interação homem-máquina é através do rastreamento de imagens, que O'hagan e Zelinsky (1997) definiram como o ramo da computação que estuda o comportamento de um objeto em uma sequência de imagens.

Devido a grande diversidade de áreas de aplicação o uso conjunto da visão computacional e processamento de imagens têm despertado o interesse de pesquisadores, acadêmicos e universidades. De acordo com Rudek (2001), tais técnicas têm sido utilizadas com sucesso em muitas aplicações relevantes, principalmente nas áreas da astronomia, medicina, análise de impressões digitais, sensoriamento remoto, multimídia, entretenimento, reconhecimento de assinaturas, manufatura, robótica de manipuladores e sistemas produtivos. Este artigo visa realizar um comparativo entre os algoritmos de segmentação de vídeos e rastreamento de objetos, e através desse comparativo demonstrar o melhor resultado combinando as duas técnicas. Para se atingir os resultados propostos foram escolhidos os algoritmos de segmentação e rastreamento: Segmentação (GONÇALVES; MONTEIRO, 2007), Segmentação utilizando Janelas Temporais (WANG; BRANDSTEIN, 1998), e a Segmentação que utiliza métodos estatísticos (STAUFFER; GRIMSON, 1999), Rastreamento por Predição (AMER, 2005), Rastreamento por Cores (MATHES; PIATER, 2005), Rastreamento utilizando filtro de Kalman (RUSSELL; NORVIG, 2003), e Rastreamento através de Fluxo Óptico (WANG et al., 2003).

1.1 Definição do Problema

O tratamento de oclusões em cenas de grande movimentação de objetos tem se tornado um forte tema de pesquisa e simultâneos testes onde se busca encontrar um

método cada vez mais qualificado. Essa considerável atividade de investigação e rastreamento de objetos a partir de sequências de vídeo tem motivado o surgimento de importantes aplicações.

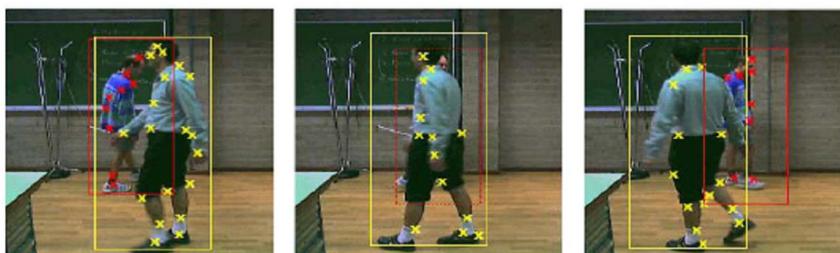
Os objetos, geralmente seres humanos, possuem comportamentos que nem sempre permitem que se faça uma previsão de qual será o seu próximo movimento, isso causa extrema dificuldade em se trabalhar com a detecção do movimento e a identificação desses objetos.

As oclusões de pessoas em cenas de vídeos, é extremamente comum quando são rastreados em cenas de multidões. Esse tratamento relacionado às oclusões é muito importante em diversas atividades, e não somente quando se fala em tecnologia da informação (TI). Pode-se também citar tudo relacionado a segurança, identificação de pessoas, rastreamento em esportes e lazer, e demais áreas onde a rastreabilidade pode ser aplicada (GABRIEL et al., 2005).

A finalidade do rastreamento de oclusões é a identificação de pessoas em movimento, e ainda, auxiliar no melhoramento de trabalhos e aplicações que são diretamente relacionadas a essa área.

Para que se possa elucidar melhor, pode-se observar na Figura 1 uma sequência de cenas nas quais o objeto é detectado e identificado. No primeiro quadro verifica-se a existência de dois objetos, o primeiro é marcado com um retângulo vermelho e o segundo com um retângulo amarelo. Já no segundo quadro ocorre um momento no qual o primeiro objeto, no caso marcado como vermelho, fica ocluído por um determinado espaço de tempo.

Figura 1 - Exemplo de cena de vídeo com objetos em movimento e oclusão



Fonte: Adaptado de Gabriel et al, 2005, p. 132.

A seguir, após o momento de oclusão, pode-se perceber que o objeto que ficou ocluído foi novamente identificado, ou seja, foi possível afirmar que o objeto selecionado inicialmente, é o mesmo objeto após o processo de oclusão.

2 Estado da Arte

Esta seção apresenta os conceitos que são necessários para melhor elucidar itens como a detecção do movimento, segmentação do movimento e o rastreamento dos objetos.

2.1 Detecção do Movimento

Softwares que realizam a análise de movimento humano, assim como os de vigilância automática, possuem objetos comuns (pessoas), que são os considerados alvos (aquilo que se quer rastrear). Os objetos encontram-se em movimento dentro de um ambiente observado (DEDEOGLU, 2004).

O processo para a detecção do movimento é dado através da segmentação do vídeo, ele é responsável por separar o que é fundo estático e o que são os objetos.

2.2 Segmentação do movimento

A primeira das etapas é a segmentação de movimento. Essa consiste na separação entre objetos que se movimentam na cena (*foreground*) e a parte estática ou fundo da cena (*background*), isso ocorre a partir de uma sequência de quadros de entrada. As imagens são detectadas pelo módulo de detecção do movimento. Já a saída, tem como resultado imagens binarizadas (SIEBEL, 2003). Nestas imagens, destacam-se as regiões (*blobs*) em movimento das regiões estáticas do ambiente. Na Figura 2 é apresentado um exemplo de segmentação de movimento.

Figura 2 - Exemplo de segmentação de movimento. (a) Quadro do vídeo original utilizado como referência para segmentação do movimento. (b) Quadro resultante do processo de segmentação.



Fonte: Pets, 2006, p. 35.

Para que seja possível realizar a etapa de segmentação de movimento a partir de um quadro de entrada, o módulo detector de movimento é mantido um modelo de fundo da cena (*background model*), isto é, um molde da aparência do ambiente sem a presença de pessoas ou objetos em movimento (SIEBEL, 2003). Nesse modelo realiza-se a aplicação de um algoritmo para classificar cada *pixel* do quadro de entrada como um *pixel* que pertence a um objeto alvo (em movimento). Os modelos utilizados podem sofrer algumas mudanças que podem afetar no desempenho, tais como climáticas, variações na iluminação do ambiente, sombras, superfícies espelhadas, objetos com movimentos oscilatórios com baixa amplitude (como galhos e folhas de vegetação sob efeito de ventos), dentre outros, assim, fazem da segmentação de movimento uma tarefa difícil, podendo comprometer a confiabilidade e o desempenho do algoritmo utilizado nessa segmentação.

De acordo com Lara (2007), não existe um algoritmo que seja suficientemente robusto para se adaptar as diversas situações apresentadas pelo ambiente. No entanto, quanto menos suposições em relação a um determinado ambiente não controlado, o algoritmo necessita de uma maior robustez. Dessa forma, deve-se ter consciência que o tipo de aplicação final para a qual a segmentação de movimento será utilizada.

2.3 Rastreamento

Realizada a etapa de detecção de movimento, os objetos alvo que foram detectados na aplicação estão separados dos elementos estáticos da cena (fundo da cena) e estarão a partir daí disponíveis para o processamento das etapas de alto nível, como rastreamento e detecção de eventos ou reconhecimento de ações, normalmente o rastreamento. A fase do rastreamento consiste em estabelecer relações temporais entre objetos alvo de quadros de vídeo consecutivos, isto é, identificar, no quadro corrente de um vídeo, a localização dos objetos alvo detectados no quadro anterior, realizando assim a diferença entre os frames.

O funcionamento do módulo de rastreamento é baseado em um algoritmo de rastreamento. Na literatura pode-se encontrar diversos critérios e algumas taxonomias para classificar algoritmos de rastreamento em sistemas de análise de movimento humano. Cita-se como exemplo, algumas discussões detalhadas sobre os tipos de algoritmos de rastreamento em Moeslund e Hilton (2006), Wang et al. (2003) e Elgammal et al. (2000).

Este artigo, baseia-se por questões de simplicidade, adotar uma classificação semelhante à utilizada por Amer (2005) e Elgammal et al. (2000), onde, os algoritmos selecionados realizam o rastreamento de forma que podem ser divididos em algoritmos que rastreiam partes do corpo das pessoas (cabeça, tronco e membros) e algoritmos que rastreiam os objetos como um todo, esse critério foi estipulado pois os vídeos escolhidos terem algumas interferências, como vento, e outros objetos que podem ser confundidos com os objetos, por estarem em movimento.

Sistemas de rastreamento que utilizam como base os algoritmos baseados em predição de posição ou estimativa de movimento vêm sendo bastante utilizados em sistemas de vigilância automática. Esses algoritmos utilizam ferramentas estatísticas como Filtros de Kalman ou Redes Bayesianas dinâmicas para estimar a posição dos objetos alvo no quadro seguinte de um vídeo.

Quando se fala em rastreamento de oclusões, pode-se citar as seguintes situações que podem ser encontradas: Novo, Rastreando, Perdido, Oclusão, Divisão e ainda objeto desconhecido.

3 Desenvolvimento

O método proposto se caracteriza por meio de um *framework* que realiza a leitura dos vídeos e aplica aos mesmos os algoritmos selecionados, conforme citados anteriormente. Esses algoritmos referem-se as etapas de segmentação, a qual consiste na separação entre objetos que se movimentam na cena (*foreground*) e a parte estática ou fundo da cena (*background*), a partir de uma sequência de quadros de entrada.

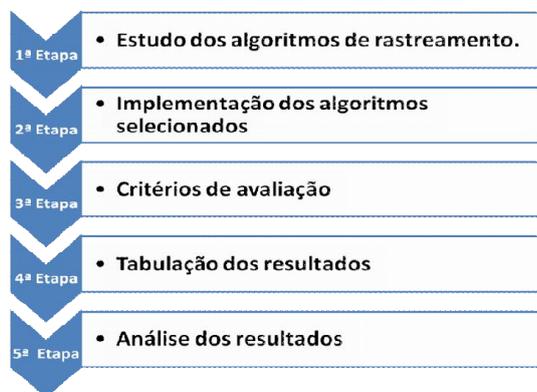
Realizado o processo de segmentação, o mesmo resulta em imagens binarizadas, onde estão os objetos em movimento, e assim, ocorre o processo de rastreamento. O rastreamento mira os objetos que são considerados alvos da aplicação, os quais se encontram separados dos elementos estáticos da cena e estarão disponíveis para o processamento das etapas de mais alto nível, como rastreamento e detecção de eventos ou reconhecimento de ações. O rastreamento consiste em estabelecer relações temporais entre objetos alvo dos quadros de vídeo consecutivos, isto é, identificar, no quadro corrente de um vídeo, a localização dos objetos alvo detectados no quadro anterior.

Assim, a proposta foca o estudo comparativo, utilizando os resultados de diversos algoritmos que trabalham com a segmentação, seleção do objeto e rastreamento do mesmo. O principal objetivo dessa comparação está em detectar se os objetos

selecionados na cena passam por períodos de oclusão. Dentre os itens avaliados, também são utilizados outros três critérios: Objetos que entram na cena, Objetos que saem da cena e o total de objetos presentes na cena. Assim que verificado, é realizado o comparativo entre os resultados desses diferentes algoritmos com a base PETS 2009 S2.L1.

Esse comparativo tem como critérios de avaliação a contagem ao longo dos quadros dos objetos *frame a frame*, a robustez de cada algoritmo, o tipo de oclusão de objetos por *frame*, os erros de contagem nas situações de oclusões, a quantidade geral de quadros e o percentual de acertos. Esses resultados gerados pelos algoritmos individualmente são comparados com a rotulação manual da base de dados.

Figura 3 - Representação do processo para solução da problemática de tratamento de oclusões de objetos em cenas de vídeo.



Fonte: Elaborada pelo autor.

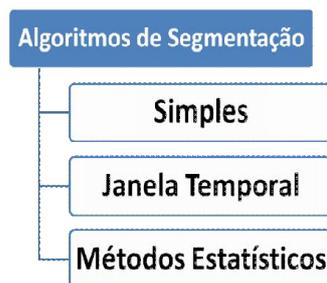
Para que se possa entender melhor o processo a ser seguido pelo método proposto, a Figura 3, ilustra as etapas a serem realizadas.

3.1 Estudo Dos Algoritmos Selecionados

A primeira etapa consiste no estudo dos algoritmos de rastreamento e segmentação.

Para este artigo, utilizou-se algoritmos de segmentação de objetos que estão sendo utilizados para esse compor o método proposto são os apresentados na Figura 4.

Figura 4 - Representação dos algoritmos de Segmentação



Fonte: Elaborada pelo autor.

A definição para a segmentação de imagens esta diretamente relacionada à área na qual será aplicada. Dentro da área de visão computacional, a segmentação refere-se ao processo de decomposição de uma imagem digital em vários segmentos (regiões) que a formam Jain, (1989).

Pode-se entender melhor o seu funcionamento a partir da descrição de cada um dos métodos de segmentação:

- a) **Segmentação simples:** A segmentação simples, também conhecida como modelo de segmentação de subtração de imagens (ou *background subtraction*) é um dos mais simples modelos existentes de segmentação de imagens. Esse modelo consiste na utilização de imagens estáticas de ambientes, desde que nesse momento não haja objetos que se encontrem em movimento na cena. Verificado isso, cada quadro do vídeo é subtraído da imagem estática, previamente definida. Assim, resultando na subtração dos frames, possibilitando diferenciar o que são objetos e o que é o fundo da cena. Em outras palavras, baseia-se apenas na subtração de um frame atual por outro antigo. Se a diferença for maior que um determinado *threshold* (limiar) o pixel é considerado como objeto, caso contrário, será considerado fundo.
- b) **Segmentação por Janela Temporal:** Os modelos de diferença temporal, ou modelos de segmentação por janela temporal, trabalham similarmente ao modelo de segmentação simples. A segmentação de diferença temporal é bastante utilizada por ser uma segmentação bastante simples. A mesma consiste em subtrair os *pixels*, *pixel a pixel*, de duas ou mais imagens de entrada consecutivas aplicadas a um limiar de classificação previamente definido.
- c) **Segmentação baseada em Métodos Estatísticos:** Os modelos que utilizam métodos estatísticos como forma de segmentação, são classificados como uma evolução dos modelos de segmentação que utilizam subtração de imagens.

Consiste da subtração *pixel a pixel* de duas ou mais imagens de entrada consecutivas aplicadas a um limiar de classificação previamente definido. Wang e Brandstein (1998) utilizaram a diferença temporal entre quadros para detectar movimento em seu sistema e rastreamento de faces em imagens de vídeo conferência. Após a classificação dos *pixels* em *background* ou *foreground*, aplicou-se um algoritmo de checagem de vizinhança para eliminar ruídos resultantes da classificação.

Outra ação dentro dessa etapa é a detecção dos objetos. Nessa etapa ocorre a seleção dos objetos de interesse. A detecção de objetos nas cenas de vídeo ocorre para que se possam manter os objetos selecionados e identificados. Quando os objetos são detectados e marcados, esses objetos recebem um valor randômico para que se possa identificá-los.

Para construção de artigo utilizou-se quatro abordagens de rastreamento, também podendo ser chamados de algoritmos de rastreamento, esses algoritmos de rastreamento são os responsáveis pelo seguimento dos objetos dentro das cenas dos vídeos. Cada um deles possui maneiras diferentes de identificar o objeto e dar início ao seguimento (*tracking*) do mesmo, por isso, faz-se necessário que seja realizada a seguir a descrição e a explicação do funcionamento de cada um deles.

Os algoritmos de rastreamento encontram-se descritos de acordo com a ordem apresentada na Figura 5, a qual exemplifica as quatro abordagens de algoritmos de rastreamento.

Figura 5 - Representação dos algoritmos de rastreamento



Fonte: Elaborada pelo autor.

Faz-se necessário que cada um desses algoritmos sejam descritos de modo claro, para que se possa entender o seu funcionamento.

- a) **Rastreamento por predição:** O rastreamento de objetos que utiliza como método a predição, o mesmo trabalha da forma a armazenar uma determinada quantidade de caminhos que os objetos pode fazer. Desse modo, realiza uma

previsão dos caminhos que podem ser percorridos pelo objeto nos frames. O algoritmo verifica o *blob* local dentro de um milite de frames, onde verifica a quantidade de pixels de onde havia feito a previsão. Assim, se o *blob* seguir algum dos caminhos previstos, o algoritmos considera como sendo o mesmo objeto, ou mesmo um outro objeto, e ainda pode considerar como um estado de oclusão.

- b) **Rastreamento baseado em cores:** Este método de rastreamento tem por base as cores dos objetos para identificar o movimento nas cenas de vídeo, realiza o rastreamento do objeto pela sua cor, porém, é extremamente comum que o algoritmo se confunda, pois é fácil encontrar mais de uma pessoa com a mesma tonalidade de roupa.
- c) **Rastreamento por Filtro de Kalman:** O filtro de Kalman é definido por conjuntos de equações matemáticas que propiciam uma implementação computacional eficiente para estimar o estado de um processo, de maneira tal que o erro inerente a essa estimativa seja minimizado (Welch e Bishop, 1995). No caso do rastreamento de objetos em vídeo, o processo em questão é a dinâmica dos objetos previamente segmentados, enquanto que o estado do processo é representado pelo conjunto de características que serão estimadas (geralmente, posição e/ou velocidade).
- d) **Rastreamento de Fluxo Óptico:** Fluxo óptico é a distribuição 2D da velocidade aparente do movimento dos padrões de intensidade no plano da imagem (Horn, 1986). Em outras palavras, o campo do fluxo óptico consiste de um campo denso de velocidade onde a cada pixel no plano da imagem está associado um único vetor de velocidade. Para fins de visualização, o campo é amostrado em uma malha e chamado de diagrama de agulhas (*needle map*). Se for conhecido o intervalo de tempo entre duas imagens consecutivas, os vetores da velocidade podem ser convertidos em vetores de deslocamento e vice-versa (Shi e Sun, 2000). A detecção de objetos se movimentando na cena, ou a câmara se movimentando e os objetos parados ou ainda a câmara e os objetos se movendo, são problemas mais difíceis em comparação com a detecção do movimento. Outros problemas mais complexos incluem a detecção do objeto (se movendo em um ambiente com vários objetos), a detecção da trajetória desses objetos e a predição das trajetórias. A análise do movimento é chamada de análise dinâmica de imagens e é baseada em um pequeno número de imagens em uma sequência,

algumas vezes duas ou três imagens. Esse caso é similar a uma análise estática de imagens e o movimento é atualizado através da correspondência entre pares de pontos de interesse ou feições/regiões nas imagens subsequentes. Uma outra abordagem analisa o movimento a partir da computação do fluxo óptico onde uma pequena diferença de tempo entre as imagens é requerida e não devem ocorrer mudanças significativas entre duas imagens consecutivas. A computação do fluxo óptico resulta na direção e velocidade dos pontos na imagem. Um objetivo imediato da análise de imagens baseada em fluxo óptico é determinar o campo de movimento. O fluxo óptico nem sempre corresponde ao verdadeiro campo de movimento por causa das mudanças de iluminação, mas representa uma aproximação dele.

3.2 Implementação dos algoritmos Segmentação e de Rastreamento

A segunda etapa consiste na implementação dos algoritmos de rastreamento selecionados e, também, dos algoritmos de segmentação e detecção de objetos nas cenas de vídeos.

Para a implementação dos algoritmos, foi utilizada a linguagem de programação C e C++, na plataforma Eclipse IDE, o sistema operacional Linux com a distribuição UBUNTU 10.4 e a biblioteca OpenCV, versão 1.0 para a leitura dos vídeos. Essa plataforma irá possibilitar que seja realizada as combinações necessárias entre os algoritmos de segmentação e rastreamento de objetos em cenas de vídeos. Essa combinação entre os algoritmos é dada pela Tabela 1 logo adiante.

3.3 Critérios de Avaliação

A terceira etapa consiste em explicar os critérios que foram utilizados para a avaliação dos algoritmos. Foi considerado como critério de avaliação dos algoritmos alguns quesitos. Esses quesitos servem como base para a rotulação dos resultados e a análise dos mesmos. Para cada vídeo interpretado pela *framework* se obtém diferentes dados, pois os algoritmos serão combinados entre os algoritmos de segmentação e de rastreamento. Ao todo foram realizados para cada vídeo o total de combinações de três algoritmos de segmentação e quatro de rastreamento, ou seja, para cada combinação o vídeo tem quatro resultados diferentes, ao total cada vídeo possui doze resultados diferentes a serem comparados. Considerando que a base rotulada da PETS 2009 S2.L1

possui cinco vídeos que foram escolhidos, foi obtido um total de sessenta resultados para serem comparados e tabulados, gerando os resultados baseados nos critérios de avaliação. Para que se possa entender melhor como ocorreu a combinação dos algoritmos de segmentação e rastreamento, segue a Tabela 1 abaixo que representa a forma que ocorrerão as combinações.

Tabela 1 - Representação do processo de combinação entre os algoritmos de segmentação e rastreamento

TABELA DE REPRESENTAÇÃO DAS COMBINAÇÕES DOS ALGORITMOS					
Algoritmos de Segmentação	01 = Simples		Algoritmos de Rastreamento	01 = Predição	
	02 = Janela Temporal			02 = B. em Cores	
	03 = Mét. Estatísticos		03 = F. de Kalman		
			04 = F. Óptico		
As combinações seguem a ordem de (Segmentação, Rastreamento)					
Combinação 01(C1)	(1,1)	(1,2)	(1,3)	(1,4)	4 Resultados
Combinação 02(C2)	(2,1)	(2,2)	(2,3)	(2,4)	4 Resultados
Combinação 03(C3)	(3,1)	(3,2)	(3,3)	(3,4)	4 Resultados
Total de resultados por vídeo →					12

Fonte: Elaborada pelo autor.

Observada a Tabela 1, percebe-se a maneira de combinação dos algoritmos, portanto, sendo que tem-se 12 resultados por vídeo, e um total de cinco vídeo, ao final se obtém o valor de 60 resultados a serem analisados.

Sobre os critérios de avaliação, os dados considerados para identificação de cada avaliação, são iguais para todos os vídeos. Essa representação é ilustrada por tabelas, onde cada tabela representa a combinação do método, ou seja, os algoritmos de segmentação de cenas de vídeo e o rastreamento de objetos nas cenas.

A avaliação dos experimentos se deu através da comparação entre os arquivos rotulados da base de dados e os arquivos gerados pelo programa, esses que são comparados linha a linha. Essa avaliação leva em consideração um algoritmo de segmentação utilizado para cada algoritmo de rastreamento, sendo os critérios avaliados:

- A quantidade de pessoas que entram na cena;
- A quantidade de pessoas que saem da cena;
- A quantidade de pessoas por cena;
- O ocorrência de oclusão e o tipo de oclusão (P = Oclusão Parcial; T = Oclusão Total; N = Não ocluso ou em transito normal);

- O cálculo gerado vezes a rotulação.

Para os quatro primeiros itens, essas quantidades de objetos é contabilizada *frame a frame*. No quinto item, o resultado gerado pela implementação é calculado com os valores da rotulação manual, gerando um percentual de acertos/erros na contagem e também um percentual no acerto e erro direcionado as oclusões de cada vídeo. O último item diz respeito ao foto de ocorrer ou não oclusão, considera-se o tipo de oclusão (Normal ou Não Ocluso = N; Parcial = P; e Total = T).

A análise dos erros é baseada nos critérios estabelecidos para a identificação da oclusão nos vídeos. Para cada vídeo é realizada a verificação dos seguintes itens:

- **Acerto / Erro:** trata dos frames rotulados manualmente com os frames gerados pelo programa. O exemplo a seguir ilustrado pela Tabela 2, onde pode-se observar a comparação entre os frames rotulados e os frames gerados pelo programa, nesse exemplo cita-se a utilização do algoritmo de segmentação de Subtração combinada com o algoritmo de rastreamento de Predição. Utilizou-se o vídeo 001.

Tabela 2- Exemplo da comparação entre frames rotulados e frames gerados.

ROTULADO	GERADO	CORRETO/ERRADO
F000(0,0,3,N);	F000(0,0,0,N);	ERRADO
F001(0,0,3,N);	F001(0,0,3,N);	CORRETO
F002(0,0,3,N);	F002(0,0,3,N);	CORRETO
F003(0,0,3,N);	F003(0,0,3,N);	CORRETO
F004(0,0,3,N);	F004(0,0,3,N);	CORRETO
F005(0,0,3,N);	F005(0,0,3,N);	CORRETO
F006(0,0,3,N);	F006(0,0,3,N);	CORRETO
F007(0,0,3,N);	F007(0,0,3,N);	CORRETO
F008(0,0,3,N);	F008(0,0,3,N);	CORRETO
F009(0,0,3,N);	F009(0,0,3,N);	CORRETO
F010(0,0,3,N);	F010(0,0,4,N);	ERRADO

Fonte: Elaborado pelo autor.

- **Tipo de oclusão:** que trata das diferentes maneira que foram utilizadas para a identificação do tipo de oclusão ocorrida em cada vídeo, seguem :

P = Parcial: Objeto ocluso parcialmente por outro objeto;

T = Total: Objeto totalmente ocluso por outro objeto;

N = Normal ou Não Ocluso: Objeto em transito normal, sem sofrer nenhum tipo de oclusão.

O resultados dos percentuais é obtido através de um mecanismo implementado, onde, compara-se o arquivo rotulado de cada vídeo com o arquivo gerado pelo

programa. Dessa forma obtém-se os resultados de cada vídeo com todas as combinações possíveis.

3.4 Tabulação Dos Resultados

A quarta etapa consiste na tabulação dos resultados, esses, são dados pelos itens que podem ser observados a seguir na Tabela 3. Essa tabela é padrão para todas as combinações dos resultados dos algoritmos de segmentação e rastreamento. Os dados obtidos são decorrentes da comparação entre a base de dados rotulada manualmente, e o arquivo gerado pelo programa.

Tabela 3 - Representação do processo de avaliação do método

TABELA COMPARATIVA DOS RESULTADOS OBSERVADOS								
STAUFFER	CORES		FLUXO OPTICO		KALMAN		PREDIÇÃO	
Vídeos	% Erro na Contagem	% Erros de Oclusão	% Erro na Contagem	% Erros de Oclusão	% Erro na Contagem	% Erros de Oclusão	% Erro na Contagem	% Erros de Oclusão
V1								
V4								
V5								
V6								
V7								
Valor Médio								

Fonte: Elaborada pelo autor.

3.5 Análise Dos Resultados

A quinta etapa trata da análise dos resultados. Após a tabulação dos mesmos é necessário que se faça a análise dos resultados proporcionados pelos algoritmos, com a base de dados já rotuladas manualmente. Nessa etapa o comparativo irá trazer resultados como a existência de oclusão no vídeo, os tipos de oclusão, quantidade de erros e acertos, e um percentual das combinações dos algoritmos de segmentação e rastreamento.

4 Resultados

Esta seção apresenta a visão geral dos resultados e também a análise individual dos principais resultados observados.

4.1 Construção dos Experimentos

Conforme Tabela 3, no item 3.3, todos os algoritmos de segmentação de imagens e de rastreamento de objetos foram combinados de forma igualitária, sendo realizada todas as comparações possíveis entre esses sete algoritmos.

Para essa comparação utilizou-se de uma base de dados rotulada manualmente. Essa base rotulada é comparada com os arquivos gerados pelo programa de comparação. Cada um desses arquivos é composto por 795 frames, onde, todos são comparados linha a linha e fornecem um percentual de acerto e erro com base nos critérios de tabulação já citados.

Das comparações realizadas, resultaram um total de 60 resultados, isso devido a utilização de cinco vídeos. Sendo que utilizou-se 3 algoritmos de segmentação e 4 algoritmos de rastreamento, tem-se 12 resultados para cada vídeo, totalizando o 60 resultados.

Embora todos os resultados possam ser considerados satisfatórios, levando em consideração os critérios estabelecidos para a avaliação dos vídeos rotulados X gerados, a seguinte Tabelas 06 apresenta todos os resultados obtidos e logo abaixo também é apresentada a discussão sobre cada resultado, fornecendo assim, um panorama geral da técnica que melhor obteve desempenho, juntamente com a análise dos erros referentes a contagem incorreta das oclusões.

4.2 Resultados Experimentais

Os resultados preliminares estão divididos na rotulação da base de testes utilizada neste trabalho, e nos resultados dos experimentos alcançados com as implementações dos algoritmos de segmentação combinados com os algoritmos de rastreamento de objetos em cenas de vídeo.

A Base escolhida para a rotulação foi a Base PETS 2009 Eleventh IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, da qual foi utilizada a Base de Monitoramento de Pessoas S2 e o cenário de S2.L1 que contém imagens de pessoas caminhando, onde os objetos são pessoas dispersas.

A seguinte Tabela 5 traz a representação dos vídeos que serão utilizados para os testes. Os vídeos foram retirados da base PETS 2009 S2.L1, que são pastas que

possuem o caminhamento de pessoas em cenas de vídeo. Os vídeos tem uma frequência de sete quadros por segundo, pode-se verificar isso na coluna número de frames original. Quando utilizado o OpenCV para juntar os vídeos, os frames sofrem uma junção, assim, o vídeo passa ter trezentos e noventa e sete frames, como é possível na coluna número de frames no vídeo, dando uma média de quatro frames por segundo. A Tabela também contempla o nome do vídeo, que foi mantido como original da base PETS 2009, a resolução de cada vídeo, e o tempo individual do vídeo. Também o tempo total dos cinco vídeos, juntamente com o total de frames original e o total de frames após a junção dos mesmos.

Tabela 5 - Representação dos vídeos utilizados.

REPRESENTAÇÃO DOS VÍDEOS UTILIZADOS – BASE PETS 2009 S2.L1				
Vídeo	Resolução (Pixeis)	Nº de frames original	Nº de Frames no vídeo	Tempo do Vídeo (mm:ss)
View_001	768 x 576	795	795	00:01:47
View_004	768 x 576	795	795	00:01:53
View_005	720 x 576	795	795	00:01:53
View_006	720 x 576	795	795	00:01:53
View_007	720 x 576	795	795	00:01:53
		Total = 3975	Total = 3975	Total = 00:09:19

Fonte: Elaborada pelo autor.

Para que seja possível identificar melhor os vídeos, a Tabela 5, a seguir mostra a identificação do vídeo e uma imagem que representa o mesmo, ilustrada na Figura 6.

Figura 6 - Representação dos vídeos utilizados.



Fonte: Elaborada pelo autor.

Todos os vídeos foram rotulados manualmente. A descrição da rotulação da base está da seguinte forma: Todos os frames foram rotulados individualmente, a ordem que

segue para cada frame, é primeiro a quantidade de pessoas que entram na cena, depois a quantidade de pessoas que saíram da cena, depois o número de pessoas que estão na cena, e por último o tipo de oclusão. O tipo de oclusão se caracteriza por Normal ou Não ocluso, representado por N, parcial, representado por P, e a oclusão total, representada por T. Exemplo: FXXX(1,0,2,N).

4.3 Análise Dos Resultados

A Tabela 6 ilustra todos os resultados encontrados, para as combinações dos algoritmos de segmentação e rastreamento para os cinco vídeos utilizados nos experimentos.

Tabela 6 - Tabela de resultados: Geral

TABELA COMPARATIVA DOS RESULTADOS OBSERVADOS								
STAUFFER	CORES		FLUXO OPTICO		KALMAN		PREDIÇÃO	
Vídeos	% Erro na Contagem	% Erros de Oclusão	% Erro na Contagem	% Erros de Oclusão	% Erro na Contagem	% Erros de Oclusão	% Erro na Contagem	% Erros de Oclusão
V1	53,00%	32,00%	51,00%	34,00%	46,00%	29,00%	49,00%	31,00%
V4	50,00%	44,00%	49,00%	42,00%	43,00%	36,00%	46,00%	40,00%
V5	40,00%	20,00%	38,00%	20,00%	32,00%	18,00%	37,00%	19,00%
V6	42,00%	22,00%	39,00%	20,00%	33,00%	17,00%	41,00%	21,00%
V7	47,00%	23,00%	48,00%	24,00%	40,00%	20,00%	44,00%	19,00%
Valor Médio	46,40%	28,20%	45,00%	28,00%	38,80%	24,00%	43,40%	26,00%
SUBTRAÇÃO	CORES		FLUXO OPTICO		KALMAN		PREDIÇÃO	
Vídeos	% Erro na Contagem	% Erros de Oclusão	% Erro na Contagem	% Erros de Oclusão	% Erro na Contagem	% Erros de Oclusão	% Erro na Contagem	% Erros de Oclusão
V1	49,00%	33,00%	48,00%	33,00%	40,00%	26,00%	45,00%	30,00%
V4	51,00%	44,00%	47,00%	41,00%	42,00%	35,00%	52,00%	45,00%
V5	45,00%	23,00%	41,00%	22,00%	36,00%	19,00%	44,00%	23,00%
V6	44,00%	22,00%	43,00%	21,00%	34,00%	16,00%	41,00%	21,00%
V7	46,00%	22,00%	43,00%	21,00%	39,00%	19,00%	45,00%	21,00%
Valor Médio	47,00%	28,80%	44,40%	27,60%	38,20%	23,00%	45,40%	28,00%
SEQUENCIA	CORES		FLUXO OPTICO		KALMAN		PREDIÇÃO	
Vídeos	% Erro na Contagem	% Erros de Oclusão	% Erro na Contagem	% Erros de Oclusão	% Erro na Contagem	% Erros de Oclusão	% Erro na Contagem	% Erros de Oclusão
V1	49,00%	22,00%	48,00%	24,00%	41,00%	14,00%	45,00%	21,00%
V4	55,00%	43,00%	51,00%	41,00%	50,00%	39,00%	52,00%	41,00%
V5	43,00%	22,00%	40,00%	19,00%	36,00%	17,00%	41,00%	21,00%
V6	48,00%	21,00%	45,00%	19,00%	37,00%	16,00%	47,00%	21,00%
V7	42,00%	20,00%	40,00%	14,00%	32,00%	11,00%	38,00%	14,00%

Valor Médio	47,40%	25,60%	44,80%	23,40%	39,20%	19,40%	44,60%	23,60%
--------------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------

Fonte: Elaborada pelo autor.

4.3.1 Melhor resultado observado: Sequencia versus Kalman

A Tabela 7, ilustra os resultados que foram obtidos utilizando a combinação do algoritmo de segmentação seqüência (Diferença Temporal), com o algoritmo de rastreamento baseado no Filtro de Kalman. Essa combinação apresentou para os cinco vídeos avaliados um total de 827 erros, estes, discriminados na tabela abaixo. Tabela 7 - Tabela de Resultados: Segmentação Seqüência versus Rastreamento Kalman

Experimento: Segmentação Seqüência /Rastreamento Kalman										
Vide os	Distribuição dos Erros									
	Oclusão Total			Oclusão Parcial			Normal/Não Ocluso			Qt. Total de Erros
	Qt.	%ET	%EG	Qt.	%ET	%EG	Qt.	%ET	%EG	Qt.
001	001	20,00%	0,64%	103	32,80%	65,61%	052	20,85%	16,30%	156
004	050	89,29%	16,03%	175	69,17%	56,09%	087	50,00%	27,88%	312
005	000	0,00%	0,00%	64	21,91%	44,60%	075	19,35%	51,80%	139
006	004	30,77%	3,05%	39	15,05%	29,77%	088	22,45%	67,18%	131
007	000	0,00%	0,00%	58	12,58%	65,17%	031	12,86%	34,83%	089
Totai s	055	70,51%	6,65%	439	27,96%	53,08%	333	22,23%	40,27%	827

Fonte: Elaborado pelo autor.

Conforme se observa na Tabela 7 coluna EG (Estatística Geral), os erros da combinação da segmentação baseada em abordagem seqüência (Diferença Temporal) com o rastreamento baseado no Filtro de Kalman estão distribuídos da seguinte forma: 6,65% referentes a casos de oclusão total, 53,08% a casos de oclusão parcial e 40,27% a outras situações (como problemas na segmentação, por exemplo). Ainda na Tabela 7 é possível observar na coluna ET (Estatística por Tipo) que a taxa de erro desta combinação para os casos de oclusão total, oclusão parcial e outras situações foi de 70,51%, 27,96% e 22,23% respectivamente.

4.3.2 Pior resultado observado: Segmentação Subtração X Rastreamento Cores

A Tabela 8, ilustra os resultados que foram obtidos utilizando a combinação do algoritmo de segmentação simples (Subtração), com o algoritmo de rastreamento

baseado em cores. Essa combinação apresentou para os cinco vídeos avaliados um total de 1148 erros, estes, discriminados na tabela abaixo.

Tabela 8 - Tabela de resultados: Segmentação Subtração versus Rastreamento Cores

Experimento: Segmentação Subtração/Rastreamento por Cor										
Vídeos	Distribuição dos Erros									
	Oclusão Total			Oclusão Parcial			Normal/Não Ocluso			Qt. Total de Erros
	Qt.	%ET	%EG	Qt.	%ET	%EG	Qt.	%ET	%EG	Qt.
001	002	50,00%	0,76%	208	99,52%	78,79%	053	16,67%	20,08%	263
004	049	85,96%	14,00%	169	65,25%	48,29%	132	97,73%	37,71%	350
005	003	75,00%	1,63%	66	23,66%	35,87%	115	34,95%	62,50%	184
006	001	6,25%	0,57%	56	23,14%	32,00%	118	32,60%	67,43%	175
007	000	0,00%	0,00%	142	37,67%	80,68%	034	14,29%	19,32%	176
Totais	055	64,71%	4,79%	641	46,93%	55,84%	452	32,85%	39,37%	1148

Fonte: Elaborada pelo autor.

Conforme se observa na Tabela 8 coluna EG (Estatística Geral), os erros da combinação da segmentação baseada em abordagem simples (Subtração) com o rastreamento baseado em Cor estão distribuídos da seguinte forma: 4,79% referentes a casos de oclusão total, 55,84% a casos de oclusão parcial e 39,37% a outras situações (como problemas na segmentação, por exemplo). Ainda na Tabela 8 é possível observar na coluna ET (Estatística por Tipo) que a taxa de erro desta combinação para os casos de oclusão total, oclusão parcial e outras situações foi de 64,71%, 46,93% e 32,37% respectivamente.

5 Conclusões

Este artigo apresentou um comparativo entre técnicas de segmentação de vídeos, detecção e principalmente rastreamento de objetos, considerando situações de oclusão nas cenas de vídeos. Esse comparativo entre diversas técnicas de algoritmos, foi possível devido a uma plataforma implementada que possibilitou a combinação entre os algoritmos e a base de vídeos rotulada manualmente.

Após a comparação dos métodos implementados, pode-se ter um panorama geral dos resultados e também uma visão detalhada do desempenho de cada um dos experimentos. A visão geral dos resultados demonstra como melhor resultado a combinação dos algoritmos de segmentação simples (subtração) com o algoritmo de

rastreamento baseado no Filtro de Kalman, visto que esta combinação apresenta o menor índice de erros na contagem, sendo encontrado o valor médio de 38,20% para o conjunto de cinco vídeos. O uso do Filtro de Kalman também se destaca com os demais algoritmos de segmentação Stauffer (Métodos Estatísticos) e Sequência (Diferença Temporal), sendo a taxa 38,80% e 39,20 respectivamente.

Na avaliação individual sobre os erros que ocasionaram as falhas na contagem, pode-se observar uma maior taxa de erros na combinação do experimento que utilizou o algoritmo de segmentação simples (Subtração) e o algoritmo de rastreamento por cores, apresentando um total de 1148 erros na contagem. Esses erros podem ser discriminados por tipo, sendo divididos em oclusão total, oclusão parcial e normal ou não ocluso, sendo as taxas apresentadas 4,79%, 55,84% e 39,37% respectivamente.

Realizando a avaliação sobre o método que apresentou o melhor resultado na contagem considerando os experimentos individualmente, a combinação do algoritmo de sequência (Diferença Temporal) e o algoritmo de rastreamento de Kalman, apresentando um total de 827 erros na contagem, esses erros podem ser discriminados por tipo, sendo divididos em oclusão total, oclusão parcial e normal ou não ocluso, sendo as taxas apresentadas 6,65%, 53,08% e 40,27% respectivamente.

Levando em consideração os estudos, implementações e experimentos realizados, pode-se assegurar que os algoritmos implementados possuem similaridades, e todas são extremamente sensíveis a oclusão. Embora, de uma forma a qual era possível fazer uma previsão, pode-se afirmar que o impacto das técnicas de detecção do movimento escolhidas, são extremamente significativas nos resultados do rastreamento dos objetos nas cenas dos vídeos.

Como trabalhos futuros, podem-se citar: (i) avaliar a possibilidade da combinação dos resultados dos algoritmos de rastreamento, buscando um melhor desempenho no rastreamento, sendo que a cada momento que um algoritmo falhar, outro entre em ação, possibilitando um melhor desempenho; (ii) realizar um estudo de outras métricas para a segmentação e rastreamento de objetos em cenas de vídeos além das já aplicadas nesse projeto, baseada na contagem de pessoas e (iii) implementação de métricas que avaliem os resultados no acompanhamento e contagem de pessoas, tal qual proposto por Ellis (2009), e realizar os testes comparativos.

6 Referências

- AMER A. Voting-based simultaneous tracking of multiple video objects. **IEEE Transaction on Circuits and Systems for Video Technology**, 15(11):1448–1462, November 2005.
- DEDEOGLU Y. **Moving object detection, tracking and classification for smart video surveillance**. Master's thesis, Bilkent University, Faculty of Engineering, August 2004.
- ELGAMMAL M., D.; HARWOOD, DAVIS L. S. Non-parametric model for background subtraction. In. **ECCV '00: Proceedings of the 6th European Conference on Computer Vision-Part II**, pages 751–767, London, UK, 2000. Springer-Verlag. ISBN 3-540-67686-4.
- ELLIS A. A. Overall Evaluation of the PETS2009 Results. **Proceedings 11th IEEE International Workshop on PET**, Miami, June 25, 2009.
- GABRIEL P. F.; VERLY J.; JUSTUS G.; PIATER H.; GENON A. **The State of the Art in Multiple Object Tracking Under Occlusion in Video Sequences**. Department of Electrical Engineering and Computer Science, University of Liège, Belgium. 2005.
- GONÇALVES W. N. and MONTEIRO, J. O. Multiple mice tracking using a combination of particle filter and k-means. In **Proceedings of Simopósio Brasileiro de Computação Gráfica e Processamento de Imagem (SIBGRAPI)**, pages 173–178, Belo Horizonte, Minas Gerais, Brasil, October 2007.
- HORN, B. K. P. **Robot Vision**. London, MIT, 1986. 509p.
- JAIN, A. K. **Fundamentals of digital image processing**. New Jersey: Prentice-Hall, Inc. 1989.
- LARA C.. Segmentação de movimento usando morfologia matemática. In **Proceedings of Simpósio Brasileiro de Computação Gráfica (SIBGRAPI)**, Belo Horizonte, Minas Gerais, Brasil, October 2007.
- MATHES T.; PIATER J., **Robust Non-Rigid Object Tracking Using Point Distribution Models**. British Machine Vision Conference, pp. 849–858, 2005.
- MOESLUND T. B.; HILTON A.. A survey of advances in vision-based human motion capture and analysis. **Computer Vision and Image Understanding**, 104(2):90–126, November 2006.
- NACIONAL DE ENGENHARIA DE PRODUÇÃO, 2001, Salvador. **XXI Encontro Nacional de Engenharia de Produção**. 2001.
- O'HAGAN, R.; ZELINSKY, A. Finger Track - **A Robust and Real-Time Gesture Interface**, **Australian Joint Conference on Artificial Intelligence**, Perth, 1997.
- RUDEK M.; COELHO L. S.; CANGIOLIERI J. O. Visão Computacional Aplicada a Sistemas Produtivos: Fundamentos e Estudo de Caso. In: **XXI ENCONTRO**
- RUSSELL S. J.; NORVIG P. **Artificial Intelligence: A Modern Approach**. Pearson Education, 2nd edition edition, 2003. ISBN 0137903952.

SHI, Y. Q.; SUN, H. **Image and video compression for multimedia engineering: fundamentals, algorithms and standards**. Boca Raton, CRC, 2000. 480p.

SIEBEL N. T. **Design and Implementation of People Tracking Algorithms for Visual Surveillance Applications**. PhD thesis, Department of Computer Science, The University of Reading, Reading, UK, March 2003.

STAUFFER L. E. W. G. Adaptive background mixture models for real-time tracking. In **Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 99)**, volume 2, pages 246–252, Fort Collins, USA, 1999.

WANG, C.; BRANDSTEIN M. A hybrid real-time face tracking system. Acoustics, Speech, and Signal Processing, 1998. **ICASSP'98. Proceedings of the 1998 IEEE International Conference on**, 6, 1998.

WANG, C.; BRANDSTEIN M.. A hybrid real-time face tracking system. Acoustics, Speech, and Signal Processing, 1998. **ICASSP'98. Proceedings of the 1998 IEEE International Conference on**, 6, 1998.

WANG, L.; WANG W. H.; TAN T. **Recent developments in human motion analysis. Pattern Recognition**, 2003. 36(3):585–601, 2003.

WANGENHEIM A. V.; COMUNELLO E.. **Visão computacional**: Seminário Introdução à Visão computacional. The Cyclops Project. PPGCCINE- UFSC. 2005. Disponível em <http://www.inf.ufsc.br/~visao/> acessado em 03/07/2009.

WELCH, G.; BISHOP G.. **An introduction to the kalman filter. Technical report**, University of North Carolina, Chapel Hill, NC, USA, 1995.