

Sistema de Monitoração Remota de Baixo Custo Baseado em Vídeo Digital

Patrick Letouzé Moreira, letouze@unb.br
 Francisco Assis de Oliveira Nascimento, assis@unb.br
 Cristiano Jacques Miosso, miosso@engineer.com
 Adson Ferreira da Rocha, adson@unb.br
 Universidade de Brasília
 Departamento de Engenharia Elétrica

Resumo— Neste artigo, é apresentada uma proposta para a implementação a baixo custo de um sistema de monitoração remota baseado em vídeo na forma digital. A transmissão de um ambiente monitorado a uma central de observação é feita em função da largura de banda do meio físico, da quantidade desejada de quadros transmitidos por segundo e da qualidade subjetiva do vídeo decodificado. O meio de transmissão é constituído por uma linha privativa onde a comunicação é realizada a 28800bps.

São investigados algoritmos de dizimação/interpolação associados à codificação JPEG. Nos resultados simulados, são apresentadas, para efeito de comparação, as técnicas desenvolvidas em relação ao método JPEG *baseline*. O sistema proposto nos parece promissor no sentido do baixo custo, de sua integração a uma rede de computadores e da qualidade dos resultados experimentais.

I. INTRODUÇÃO

Sistemas de monitoração remota por imagens baseiam-se na transmissão dos sinais de um ambiente monitorado a uma central de observação, com aplicações que variam desde a segurança bancária até a monitoração de crianças e pacientes em estado grave. Nesta situação, busca-se atender a um compromisso especificado entre o número de quadros transmitidos por segundo e a qualidade das imagens.

Este artigo trata da implementação e da avaliação de desempenho de cinco técnicas de dizimação/interpolação associadas à codificação JPEG [1][2], para atender a sistemas de monitoração remota via linha privativa. Os algoritmos foram implementados visando um aumento do número de quadros transmitidos por segundo. Assim, aborda-se um sistema de segurança composto por dois módulos, o de transmissão e o de recepção. O primeiro utiliza uma *WebCam* e um computador para adquirir e processar as imagens, e um MODEM para transmiti-las. Já o segundo necessita de um MODEM para receber as imagens e de um computador para reconstituí-las.

Nos sistemas de segurança que utilizam imagens, a quantidade de quadros transmitidos por segundo deve ser suficiente para garantir a integridade da localidade sob vigilância. Neste sentido, uma taxa de atualização das imagens recebidas na ordem de 1 quadro por segundo é considerada aceitável, conforme se observa nas especificações técnicas de vários sistemas disponíveis no mercado. O *EarthCam Internet Security System (ECS1700)*[3], por exemplo, promete a atualização em menos de um segundo, dependendo da velocidade de conexão da rede. Já o

AXIS 2100 Security Cam[4] permite uma taxa de transmissão máxima por MODEM de 0,11 quadro por segundo, a 28800bps e para imagens RGB de 480 linhas por 640 colunas. Como a taxa de transmissão de imagens em níveis de cinza é no máximo 3 vezes superior à de imagens RGB (caso em que não há subamostragem das crominâncias) [2], isto resulta numa taxa não superior a 0,33 quadro por segundo para as imagens analisadas neste artigo.

Uma técnica bastante utilizada para se atingir taxas desta ordem consiste em aplicar um codificador JPEG antes da transmissão de imagens dinâmicas, e um decodificador durante a recepção. Trata-se, de fato, da técnica utilizada pelos dois sistemas comerciais mencionados.

Na seção subsequente são propostas diferentes técnicas de reamostragem/codificação a serem utilizadas nos módulos de transmissão e de recepção. Conforme será mostrado na seção III, estas técnicas permitiram uma taxa de transmissão de quadros superior à do JPEG *baseline*, com uma qualidade objetiva maior. Também se observou uma redução do efeito de bloco causado pela transformada de cossenos discreta (DCT)[5][6] aplicada pelo JPEG. Finalmente, a seção IV apresenta as conclusões dessa análise bem como da comparação entre as diferentes técnicas.

II. DESCRIÇÃO DOS ALGORITMOS DE COMPRESSÃO PROPOSTOS

A. Visão Geral

Os algoritmos propostos aplicam-se a imagens em níveis de cinza, podendo ser generalizados para imagens coloridas quando aplicados a cada um dos componentes R,G e B da imagem separadamente. Foram utilizadas imagens de 480 linhas por 640 colunas com 256 níveis de cinza.

Tanto a codificação quanto a decodificação foram realizadas em duas etapas. A primeira, constituída da dizimação da imagem original seguida da codificação JPEG *baseline*, e a segunda, constituída da decodificação JPEG *baseline* e a subsequente interpolação da imagem.

Foram avaliados quatro tipos de vizinhança de *pixels*: a vizinhança-de-4 e a vizinhança-de-8 [7], e uma variação de cada uma delas pela inclusão do próprio *pixel* em questão. Como o próprio *pixel* não é conhecido de antemão, a reconstituição neste caso é feita em duas iterações. Por outro lado, os *pixels* nos limites superior e inferior, esquerdo e

direito da imagem, possuem a vizinhança incompleta. Estas regiões são tratadas de forma particular dependendo da técnica utilizada, conforme será abordado na seção II-B.

Verifica-se que a compressão depende do tipo de ditzimação escolhido. Foram avaliadas cinco técnicas diferentes, caracterizadas pela disposição geométrica dos *pixels* ditzimados. A próxima seção as descreve detalhadamente.

B. Técnicas implementadas

As diferentes possibilidades de disposição geométrica dos *pixels* eliminados durante a fase de codificação conduziram a várias técnicas, cujo desempenho é avaliado neste artigo.

Na primeira delas, eliminam-se todos os *pixels* nas linhas pares da imagem (figura 1), reduzindo em 50% o tamanho da imagem a ser processada pelo codificador JPEG. Esta técnica foi designada *extlin2* (extração das linhas pares).

Para descompressão da imagem, as linhas pares descartadas devem então ser estimadas a partir das linhas ímpares, disponíveis após a aplicação do decodificador JPEG. Numa primeira etapa, cada *pixel* naquelas linhas é obtido pela interpolação de seus vizinhos superior e inferior, o que fornece uma primeira estimativa da informação a ser reconstituída (figura 2). Um processo iterativo é então aplicado até que ocorra a convergência dos *pixels* recuperados, que em cada iteração são reavaliados como a média de todos os *pixels* numa vizinhança pré-definida (figura 3). Note-se que alguns *pixels* destas vizinhanças são também estimados, o que justifica o uso de um processo iterativo.

No entanto, os *pixels* a serem reconstituídos nas fronteiras da imagem (limites inferior e superior, esquerdo e direito) exigem um tratamento particular durante a reconstituição, visto serem incompletas suas vizinhanças. Este problema foi contornado pela duplicação das linhas e colunas extremas da imagem, antes das iterações. Terminada a reconstituição, estas linhas são então eliminadas.

Analogamente à eliminação de linhas, pode-se eliminar todos os *pixels* nas colunas pares, novamente obtendo uma redução de 50% da imagem. A implementação desta técnica, designada *extcol2*, é realizada pela aplicação da técnica anterior à transposta da imagem original, sendo necessária também a transposição da imagem reconstituída.

Outra técnica implementada foi a eliminação das linhas e colunas pares, sendo obtida pela aplicação em seqüência das duas anteriores. Neste caso, são eliminados 75% dos dados da imagem. Este procedimento foi chamado de *extlico2*.

A técnica seguinte consiste na eliminação dos *pixels* num padrão semelhante a um tabuleiro de xadrez, reduzindo o tamanho da imagem em 50% (*elxadrez*) (figura 4-a). A ditzimação neste formato pode ser realizada pela exclusão das linhas pares após o deslocamento das colunas pares para cima, sendo o *pixel* do limite superior deslocado para o limite inferior (figuras 4-b e 4-c). Em seguida se aplica a codificação JPEG. O processo inverso prepara a imagem para ser reconstituída, ou seja, após a decodificação JPEG faz-se o deslocamento das colunas pares para baixo com o *pixel* do limite inferior deslocado para o limite superior (figura 5-a). É então realizada a estimativa dos *pixels*.

Num primeiro momento, realiza-se a interpolação a partir dos *pixels* preservados, num total de 2 ou 3 nos limites da imagem e de 4 nas demais regiões (figura 5-b). Em seguida, inicia-se o processo iterativo em que são atualizados os valores dos *pixels* reconstituídos pela média de todos os vizinhos, de forma inteiramente análoga às técnicas anteriores (figuras 5-b e 5-c).

O último procedimento avaliado corresponde à eliminação de *pixels* também em padrão xadrez, porém com um maior espaçamento, conforme ilustra a figura 6-a (*elxadrez2*). Isto equivale a eliminar inicialmente as linhas pares e em seguida os padrões hexagonais descritos anteriormente (figuras 6-b e 6-c). São eliminados, portanto, 75% dos *pixels* da imagem.

A decodificação é neste caso realizada aplicando-se o mesmo procedimento descrito na técnica *elxadrez* (figura 7-a) seguido da reconstituição das linhas pares abordada na técnica *extlin2*, conforme as figuras 7-b e 7-c.

C. Comparação das Técnicas

A comparação das técnicas é realizada tendo em vista a relação sinal-ruído na imagem codificada e o tamanho do arquivo que a armazena em cada caso. Inicialmente, cada técnica é aplicada a uma mesma seqüência de vídeo, durante a fase de pré-processamento. Em seguida, é realizada uma compressão JPEG nas seqüências de vídeo obtidas, com fatores de qualidade que variam de 1 a 100. Verifica-se então qual técnica permite uma maior compressão com uma melhor relação sinal-ruído (RSR), e para qual fator de qualidade isto se verifica.

Para avaliação da RSR obtida utilizou-se a expressão [8]

$$RSR_{dB} = 10 \cdot \log \left[\frac{Var(f)}{Var(\hat{f} - f)} \right], \quad (1)$$

onde f e \hat{f} são, respectivamente, a imagem original e a reconstituída, e $Var(I)$ denota a variância do sinal I . A escolha deste critério de erro se justifica pela normalização realizada em relação à variância da imagem original ($Var(f)$). Assim, a medida de erro obtida permite comparar o desempenho das diferentes técnicas mesmo se aplicadas a imagens com diferentes níveis de energia.

Portanto, busca-se estabelecer a faixa de valores do fator de qualidade que otimiza o desempenho para cada técnica, bem como determinar se a relação sinal-ruído obtida ao longo desta faixa é satisfatória para a aplicação visada.

III. RESULTADOS DE SIMULAÇÃO

As técnicas de codificação descritas na seção II-B foram comparadas entre si e com o codificador JPEG *baseline*, tendo em vista a relação sinal-ruído dos vídeos codificados (equação 1) e o tamanho dos arquivos que as armazenam.

O gráfico na figura 8 apresenta a relação sinal-ruído obtida pelas diferentes técnicas de codificação avaliadas, em função do fator de qualidade empregado. O que se observa é que o JPEG *baseline* conduz a uma relação superior à de todas as demais técnicas, para um mesmo fator de qualidade. Este resultado já era esperado, uma vez que

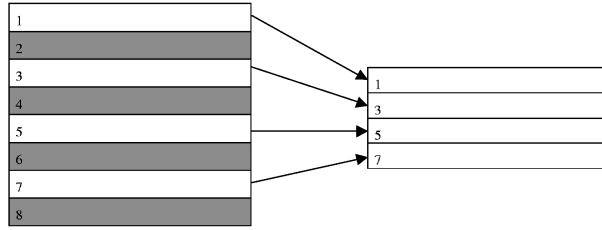


Fig. 1. Esquema da codificação de imagens pela técnica *extlin2* (eliminação das colunas pares).

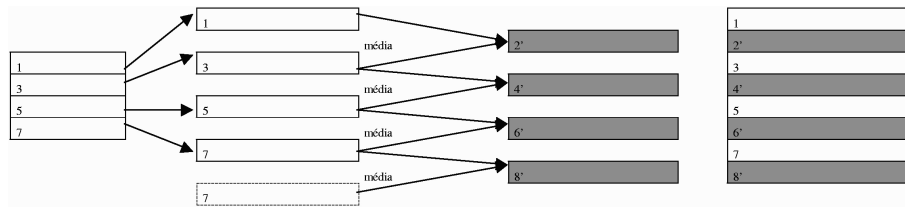


Fig. 2. Primeira etapa da decodificação de imagens pela técnica *extlin2* (estimativa das linhas pares pela média das linhas vizinhas).

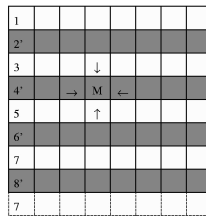


Fig. 3. Estimativa utilizando a média pela vizinhança-de-4.

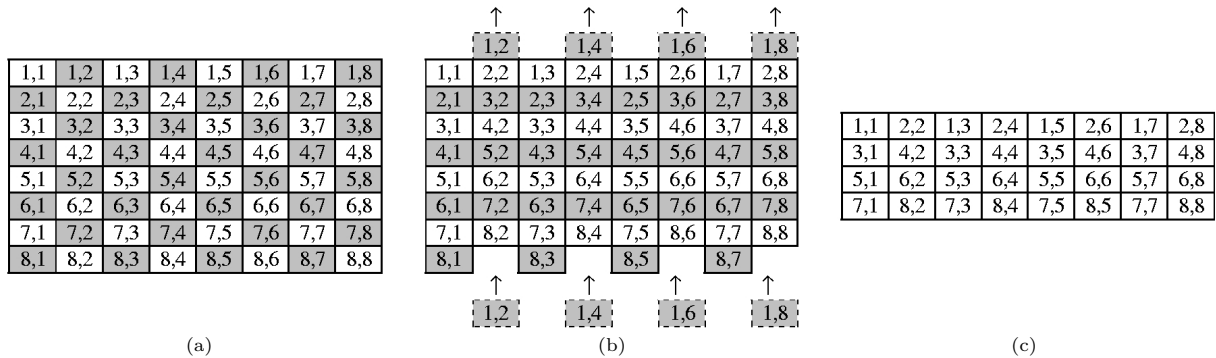


Fig. 4. Esquema da codificação de imagens pela técnica *elxadrez*.

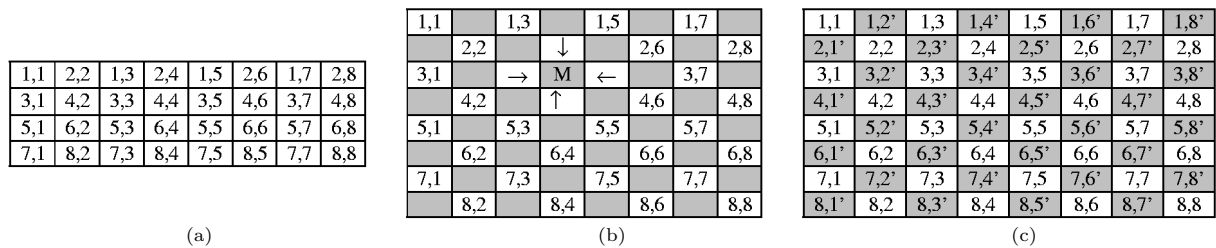
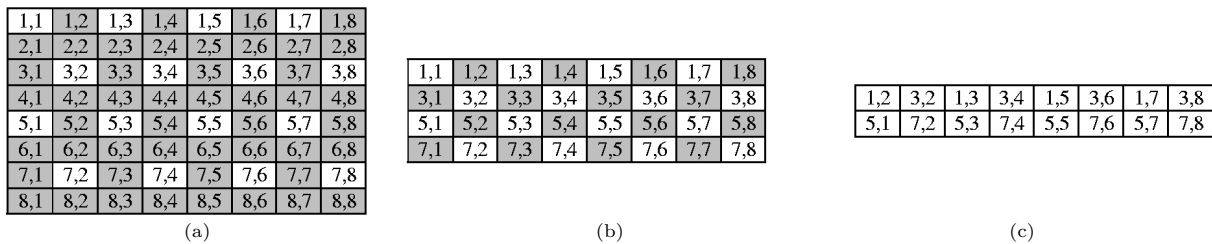
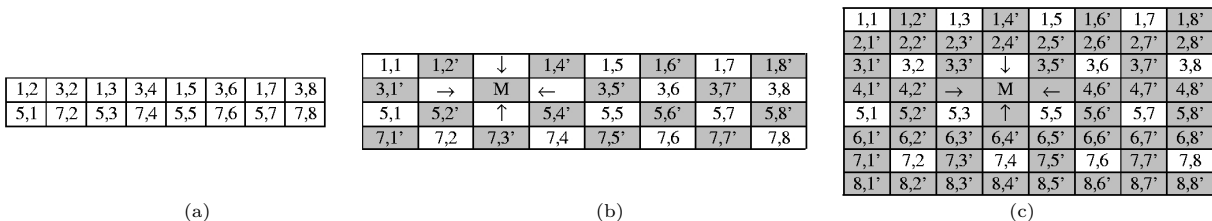


Fig. 5. Esquema da decodificação de imagens pela técnica *elxadrez*.

Fig. 6. Esquema da codificação de imagens pela técnica *exlinxa*.Fig. 7. Esquema da decodificação de imagens pela técnica *exlinxa*.

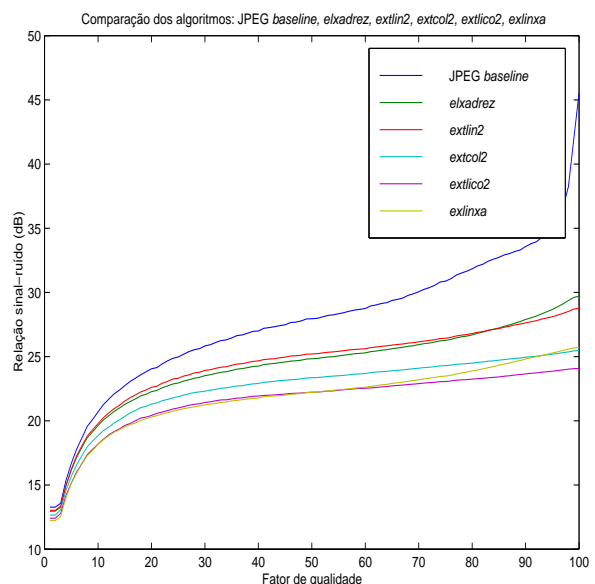
as técnicas propostas se baseiam na dizimação da imagem previamente à aplicação do codificador JPEG.

Cabe ressaltar, no entanto, que o interesse não é obter técnicas de codificação que melhorem a relação sinal-ruído para um mesmo fator de qualidade, mas sim que aumentem esta relação para um mesmo valor da razão *bits/pixels*, o que em última análise significaria a capacidade de armazenar imagens com melhor qualidade em arquivos menores. De fato, é mostrado no gráfico 9 que a imagem codificada pelo JPEG *baseline* é sempre armazenada em arquivos maiores do que os gerados por todas as outras técnicas, para todos os fatores de qualidade. Isto significa que a relação sinal-ruído superior é obtida às custas de uma maior razão *bits/pixels*.

O que as técnicas aqui apresentadas permitiram foi justamente obter uma RSR maior para uma tamanho de arquivo menor, utilizando diferentes fatores de qualidade. Esses resultados são apresentados no gráfico da figura 10, onde cada curva representa a RSR pela razão *bits/pixel* de cada técnica. Observa-se que a partir de determinado ponto, quando se compara individualmente cada técnica ao JPEG, esta passa a resultar numa RSR maior, embora necessite de um arquivo menor para armazenar a imagem. Esse ponto crítico para cada técnica é apresentado na tabela I.

Por outro lado, comparando-se todos os exemplos e os métodos simultaneamente, constata-se que a técnica *extlico2*, a certa altura, apresentou uma RSR maior com uma menor razão *bits/pixel*. A tabela II destaca o momento em que o *extlico2* passa a apresentar a característica desejada.

Escolhendo o fator de qualidade 40 para o *extlico2* a partir da tabela II, foi gerada uma seqüência de cinco imagens (figura 11), onde a coluna da esquerda apresenta as imagens originais, a coluna central mostra as imagens correspondentes codificadas pelo JPEG *baseline* com fator de qualidade 12 e a terceira coluna consiste dessas imagens codificadas pela *extlico2* com fator de qualidade 40. A comparação da seqüência de imagens entre todas as técnicas está na tabela III. Observou-se que também a qualidade subjetiva das imagens codificadas por *extlico2* apresentou-se superior.

Fig. 8. Gráfico da relação sinal-ruído em dB pelo fator de qualidade, para as técnicas de codificação JPEG, *extlin2*, *extcol2*, *extlico2*, *elxadrez*, *exlinxa*.

	FQ	RSR	Razão <i>bits/pixel</i>
JPEG	31	25.92	0,2696
EXTLIN2	67	25.99	0,2663
JPEG	14	22.39	0,1829
EXTCOL2	35	22.64	0,1827
JPEG	28	25.54	0,2562
ELXADREZ	64	25.54	0,2545
JPEG	16	23.07	0,1935
EXTLICO2	76	23.10	0,1929
JPEG	17	23.32	0,1998
EXLINXA	76	23.55	0,1977

TABELA I

COMPARAÇÃO ENTRE O JPEG E AS DEMAIS TÉCNICAS DE CODIFICAÇÃO PARA O PRIMEIRO QUADRO DA SEQÜÊNCIA DE VÍDEO.

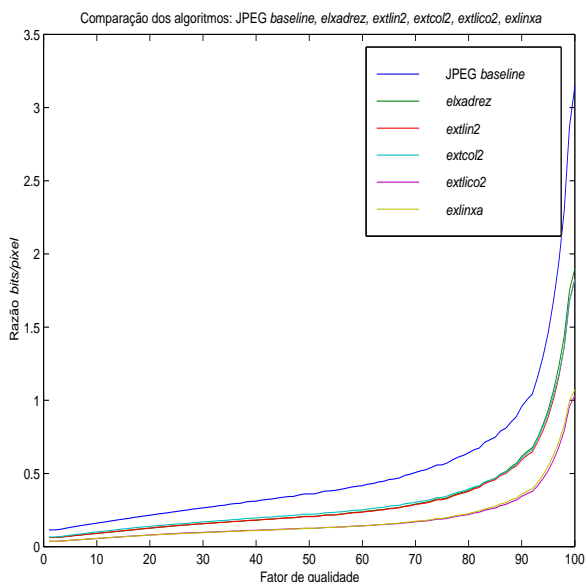


Fig. 9. Gráfico da razão $bits/pixels$ pelo fator de qualidade, para as técnicas de codificação JPEG, *extlin2*, *extcol2*, *extlico2*, *elxadrez*, *exlinxa*.

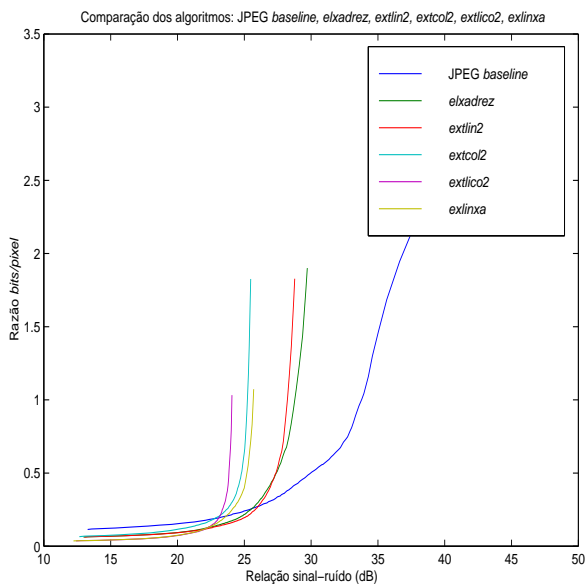


Fig. 10. Gráfico da razão $bits/pixels$ pela relação sinal-ruído, para as técnicas de codificação JPEG, *extlin2*, *extcol2*, *extlico2*, *elxadrez*, *exlinxa*.

	FQ	RSR	Razão $bits/pixel$
EXTLICO2	40	21,94	0,1116
JPEG	12	21,68	0,1717
EXTLIN2	16	21,80	0,1127
EXTCOL2	25	21,89	0,1538
ELXADREZ	17	21,88	0,1152
EXLINXA	42	21,91	0,1134

TABELA II

COMPARAÇÃO ENTRE O *extlico2* E AS DEMAIS TÉCNICAS DE CODIFICAÇÃO PARA O PRIMEIRO QUADRO DA SEQUÊNCIA DE VÍDEO.

	FQ	RSR por quadro	$bits/pixel$ por quadro	Quadros/s a 28800bps*
<i>extlico2</i>	40	21,70	0,1113	0,82
<i>JPEG</i>	12	21,65	0,1712	0,53
<i>extlin2</i>	16	21,67	0,1121	0,79
<i>extcol2</i>	25	21,79	0,1527	0,58
<i>elxadrez</i>	17	21,51	0,1148	0,78
<i>exlinxa</i>	42	21,46	0,1136	0,79

TABELA III

COMPARAÇÃO ENTRE O *extlico2* E AS DEMAIS TÉCNICAS PARA A SEQUÊNCIA DE VÍDEO. (*INCLUINDO CABEÇALHO DE 5%).

IV. CONCLUSÕES

Neste artigo, foram apresentadas cinco técnicas de amostragem/codificação de imagens que objetivam atender a um sistema de segurança baseado na transmissão de imagens dinâmicas via linha privativa.

A compressão nestas técnicas é realizada pela dizimação da imagem (*pixels* em uma configuração geométrica específica) seguida da aplicação de um codificador JPEG *baseline*. A recuperação da imagem, por sua vez, é feita utilizando-se um decodificador JPEG *baseline* e, em seguida, estimando-se a informação via técnicas interpolativas.

Essas técnicas apresentaram desempenho superior à simples aplicação do JPEG *baseline* para uma faixa particular de fatores de qualidade, uma vez que alcançaram uma maior compactação (menor razão $bits/pixel$) com uma qualidade objetiva maior (maior relação sinal-ruído).

Dentre essas técnicas, escolheu-se a *extlico2*. Esta proporcionou uma maior transmissão de quadros por segundo com uma melhor relação sinal-ruído, também indicando uma qualidade subjetiva melhor do que o JPEG *baseline*.

Os algoritmos propostos não fazem restrição quanto ao tipo de transformada utilizado, e portanto podem, em princípio, ser aplicados a qualquer codificação por transformada. Desta forma, cabe ainda avaliar seu uso com técnicas de codificação que empregam outras transformadas, como por exemplo o JPEG2000, fundamentado nas transformadas *wavelets* em substituição à DCT no JPEG.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] JPEG, *ITU-T Recommendation T.81 — ISO/IEC 10918-1:1994, Digital compression and coding of continuous-tone still images: Requirements and Guidelines*, ITU-T, 1994.
- [2] J. Miano, *Compressed Image File Formats: JPEG, PNG, GIF, XBM, BMP*, Addison Wesley Longman, Inc., July 1999.
- [3] EarthCam, *EarthCam Camera Security System ECS 1700*, <http://www.webcamstore.com/solutions/systems/index.cfm/EarthCam+Internet+Camera+Security+System/Specs>, 2001.
- [4] AXIS Communications, *AXIS 2100 Camera*, http://www.axis.com/products/cam_2100/2100to.htm#perf, 2001.
- [5] K. R. Rao e J. J. Hwang, *Techniques & Standards for Image, Video & Audio Coding*, Prentice Hall, Inc., 1996.
- [6] R. J. Clarke, *Digital Compression of Still Images and Video*, Signal Processing and its Applications. Academic Press Limited, segunda impressão, 1996.
- [7] R. C. Gonzalez e R. E. Woods, *Processamento de Imagens Digitais*, Editora Edgard Blücher LTDA., 2000.
- [8] J. S. Lim, *Two-dimensional Signal and Image Processing*, Prentice-Hall, 1990.



(a1)



(b1)



(c1)



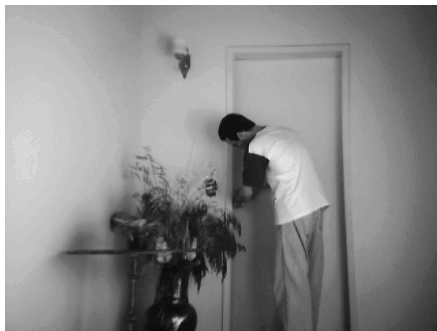
(a2)



(b2)



(c2)



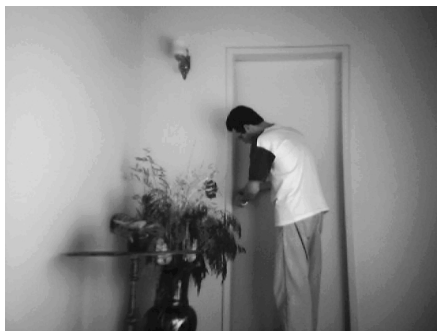
(a3)



(b3)



(c3)



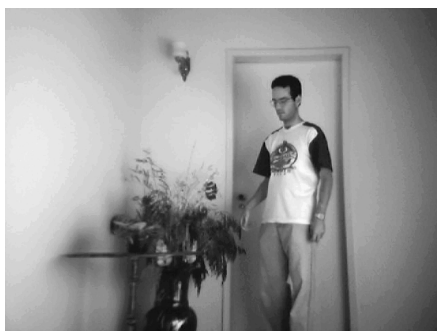
(a4)



(b4)



(c4)



(a5)



(b5)



(c5)

Fig. 11. Seqüência de imagens originais (a), codificadas pelo JPEG (b) e codificadas pela técnica *extlico2* (c).