



**UNIVERSIDADE FEDERAL DE GOIÁS**

**INSTITUTO DE QUÍMICA**

**CLASSIFICAÇÃO DE CERVEJAS POR ANÁLISE DE IMAGENS E PCA**

**Thiago César de Oliveira Silva**

Dissertação apresentada ao Instituto de Química da Universidade Federal de Goiás como exigência parcial, para a obtenção do título de Mestre em Química

Orientador: Prof. Dr. Anselmo Elcana de Oliveira

Goiânia

Junho - 2008

## **DEDICATÓRIA**

Dedico este trabalho aos meus pais, que nunca mediram esforços para minha formação intelectual, estando sempre ao meu lado durante as dificuldades encontradas em todas as etapas de minha formação.

Aos meus amigos e minha esposa, Juliana, que sempre me motivaram a continuar o desenvolvimento do trabalho, e finalmente aos colegas do laboratório, Mariana, Nilson, Deangeles e Thiago Diamond, que de alguma forma tornaram possível a conclusão deste.

## **AGRADECIMENTOS**

A Deus por me conceder o Dom da vida.

Aos meus pais, que sempre estiveram ao meu lado e me apoiaram em todos os momentos.

A minha esposa, que teve a compreensão de me ceder aos estudos mesmo após o recente enlace.

Ao meu orientador Prof. Dr. Anselmo Elcana de Oliveira, por toda sua paciência até nos momentos em que nem eu acreditava na possibilidade da conclusão do trabalho e que me mostrou ser possível o impossível.

A todos os colegas do Laboratório de Química Teórica e Computacional do Instituto de Química da Universidade Federal de Goiás (UFG), pelo apoio, incentivo e consumo das amostras durante a realização deste trabalho.

Ao Departamento de Polícia Federal em especial aos colegas da Delegacia de Polícia Federal em Presidente Prudente/SP, por permitir que eu pudesse concluir este trabalho.

A amiga, sempre presente, Mariana, que me auxiliou e incentivou nestes quase três anos.

Aos órgãos financiadores UFG, FUNAPE e CAPES.

A todas as pessoas que, de alguma forma colaboraram para a realização deste trabalho.

“O sucesso é construído à noite!

Durante o dia você faz o que todos fazem. Não conheço ninguém que conseguiu realizar seu sonho, sem sacrificar feriados e domingos pelo menos uma centena de vezes. Da mesma forma, se você quiser construir uma relação amigável com seus filhos, terá que se dedicar a isso, superar o cansaço, arrumar tempo para ficar com eles, deixar de lado o orgulho e o comodismo. Se quiser um casamento gratificante, terá que investir tempo, energia e sentimentos nesse objetivo. O sucesso é construído à noite! Durante o dia você faz o que todos fazem. Mas, para obter resultado diferente da maioria, você tem que ser especial. Se fizer igual a todo mundo, obterá os mesmos resultados. Não compare à maioria, pois infelizmente ela não é modelo de sucesso. Se você quiser atingir uma meta especial, terá que estudar no horário em que os outros estão tomando chope com batatas fritas. Terá de planejar, enquanto os outros permanecem à frente da televisão. Terá de trabalhar enquanto os outros tomam sol à beira da piscina.

A realização de um sonho depende de dedicação. Há muita gente que espera que o sonho se realize por magia. Mas toda magia é ilusão. A ilusão não tira ninguém de onde está. Ilusão é combustível de perdedores. Quem quer fazer alguma coisa, encontra um meio. Quem não quer fazer nada, encontra uma desculpa.”

*Roberto Shinyashiki*

“Corre a lenda no mundo da física e da engenharia que o besouro não pode voar. Analisando sua construção, consta que ele é pesado demais, que suas asas são muito pequenas, que seu abdomen é disforme e que o cérebro é incapaz de enfrentar as complexas tarefas do vôo. Portanto, teoricamente, o besouro não pode voar. Mas voa!”

Autor Desconhecido

# SUMÁRIO

LISTA DE FIGURAS .....	i
LISTA DE TABELAS .....	iv
LISTA DE ABREVIATURAS E SIGLAS .....	v
LISTA DE SÍMBOLOS .....	vi
RESUMO .....	vii
ABSTRACT .....	viii
1 INTRODUÇÃO .....	01
1.1 Objetivos .....	05
2 FUNDAMENTOS TEÓRICOS .....	06
2.1 Cerveja .....	06
2.2 Imagens Digitais .....	13
2.3 Análise de Componentes Principais .....	16
2.4 Análise de Imagens Digitais .....	22
3 METODOLOGIA .....	24
3.1 Amostragem .....	24
3.2 Equipamentos .....	24
3.3 Procedimento Experimental .....	25
3.3.1 Captura das Imagens de Cervejas .....	25
3.3.2 Análise das Imagens Digitalizadas .....	26
3.3.3 Análise de Componentes Principais da Imagens Digitalizadas .....	27
4 RESULTADOS E DISCUSSÃO .....	28
4.1 Análises de Imagens .....	28
4.2 Análise de Imagens em 300 ppp formato JPEG .....	35
4.3 Análise de Imagens em 600 ppp formato JPEG .....	41
4.4 Análise de Imagens em 300 ppp e 600 ppp formato JPEG .....	44
4.5 Análise de Imagens em 300 ppp formato JPEG avaliação de alteração de cor .....	46
4.6 Análise de Imagens em 300 ppp formato BMP .....	52
5 CONCLUSÕES .....	59
6 ANEXOS .....	61
7 REFERÊNCIAS BIBLIOGRÁFICAS .....	65

## LISTA DE FIGURAS

FIGURA 1	- Cubo de tons dos canais RGB.	03
FIGURA 2	- Representação da matriz de dados <b>X</b> decomposta em produto de matrizes de posto igual a 1.	16
FIGURA 3	- Uma componente principal de duas variáveis: (A) <i>loadings</i> : ângulos do vetor direção; (B) <i>scores</i> : projeções das amostras (1 a 6) na direção da componente principal. Dados são centrados na média.	16
FIGURA 4	- Fluxograma para captura de imagens de cervejas.	26
FIGURA 5	- Fluxograma para obtenção da matriz <b>X</b> .	27
FIGURA 6	- <i>Scanner</i> e placa de petri utilizados para análises adaptados com anteparos brancos	29
FIGURA 7	- Imagem de cerveja contendo a região retangular definida para análise.	30
FIGURA 8	- Imagem representativas da marca 1 em diferentes intervalos de tempo (0 h, 1 h, 2 h, 3 h, 4 h e 24 h) quando exposta ao ar	32
FIGURA 9	- Gráfico dos <i>scores</i> das duas primeiras componentes principais, PC1 e PC2, que explicam, respectivamente, 44,16% e 29,17% das variâncias dos canais RGB com 73,33% da variância acumulada das marcas 1 a 10 – resolução 300 ppp, no formato JPEG.	35
FIGURA 10	Imagens representativas das 10 diferentes amostras de cervejas com resolução de 300 ppp. Imagens das marcas 1 a 10 no formato JPEG.	36
FIGURA 11	Histograma médio das freqüências versus tons RGB das amostras de cervejas das marcas 1 a 10, com resolução de 300 ppp no formato JPEG.	36
FIGURA 11.1	Histograma médio ampliado das freqüências versus tons R das amostras de cervejas das marcas 1 a 10, com resolução de 300 ppp no formato JPEG.	37
FIGURA 11.2	Histograma médio ampliado das freqüências versus tons G das amostras de cervejas das marcas 1 a 10, com resolução de 300 ppp no formato JPEG.	37
FIGURA 11.3	Histograma médio ampliado das freqüências versus tons B das amostras de cervejas das marcas 1 a 10, com resolução de 300 ppp no formato JPEG.	38
FIGURA 12	Gráfico dos <i>loadings</i> de PC1 e PC2 das marcas 1 a 10 de cervejas. Imagens digitalizadas com resolução de 300 ppp, no formato JPEG. As numerações referem-se aos valores dos índices de cor para os canais R, G e B justapostos.	39
FIGURA 13	Gráfico dos <i>scores</i> PC1 e PC2 que explicam, respectivamente, 50,53% e 29,82% das variâncias dos canais RGB com 80,35% da variância acumulada das marcas 1 a 10, com resolução 600 ppp no formato JPEG.	40

FIGURA 14	Imagens representativas das 10 diferentes amostras de cervejas com resolução 600 ppp no formato JPEG.	41
FIGURA 15	Histograma médio das frequências versus tons Canal B de amostras das marcas 1 a 10 de cervejas, com resolução de 600 ppp no formato JPEG	42
FIGURA 16	Gráfico dos <i>loadings</i> de PC1 e PC2 das marcas 1 a 10 de cervejas resolução de 600ppp, no formato JPEG. As numerações referem-se aos valores dos índices de cor para os três canais justapostos.	43
FIGURA 17	Gráfico dos <i>scores</i> PC1 e PC2 que explicam respectivamente 48,46 % e 23,49 % da variância dos canais RGB com 71,95% da variância acumulada dos dados das 10 diferentes marcas em escalas de 300 ppp e 600 ppp no formato JPEG. A linha tracejada é indicativo da separação das imagens com diferentes resoluções à exceção das marcas 2 (12).	45
FIGURA 18	Gráfico dos <i>loadings</i> de PC1 e PC2 das marcas 1 a 10 de cervejas resolução de 300 ppp e 600 ppp, no formato JPEG. As numerações referem-se aos valores dos índices de cor para o canal B com índice dos canais justapostos.	46
FIGURA 19	Gráfico dos <i>scores</i> PC1 e PC2 que explicam 98,98% da variância acumulada dos dados da marca 1 para diferentes intervalos 1- 0 h, 2- 1 h, 3- 2 h ,4- 3 h, 5- 4 h e 6- 24 h com resolução de 300 ppp.	47
FIGURA 20	Imagens representativas da marca 1 em diferentes intervalos (0h, 1h, 2h, 3h, 4h e 24h) quando exposta à luz e ao ar. Imagens com resolução de 300 ppp no formato JPEG.	48
FIGURA 21	Gráficos de frequências x Tons RGB de amostra da marca 1 em diferentes intervalos de tempo com resolução de 300 ppp no formato JPEG.	49
FIGURA 21.1	Histograma médio ampliado de frequências x Tons R de amostra da marca 1 em diferentes intervalos de tempo com resolução de 300 ppp no formato JPEG.	50
FIGURA 21.2	Histograma médio ampliado de frequências x Tons G de amostra da marca 1 em diferentes intervalos de tempo com resolução de 300 ppp no formato JPEG.	50
FIGURA 21.3	Histograma médio ampliado de frequências x Tons B de amostra da marca 1 em diferentes intervalos de tempo com resolução de 300 ppp no formato JPEG.	51
FIGURA 22	Gráfico dos <i>loadings</i> das duas primeiras componentes principais, PC1 e PC2 para a amostra da marca 1 com resolução de 300 ppp medidas no período de 0 h, 1 h, 2 h, 3 h, 4 h e 24 h. As numerações referem-se aos valores dos índices de cor para os três canais justapostos	52

- FIGURA 23 Gráfico dos *scores* PC1 e PC2 que explicam respectivamente 66,05 % e 25,50 % de variância dos canais RGB com 91,55% da variância acumulada dos dados das 8 diferentes marcas com resolução de 300 ppp no formato BMP. 53
- FIGURA 24 Histograma médio das freqüências versus tons RGB de amostras de 8 marcas cervejas, com resolução de 300 ppp no formato BPM. 54
- FIGURA 24.1 Histograma médio ampliado das freqüências versus tons R de amostras de 8 marcas cervejas, com resolução de 300 ppp no formato BPM. 55
- FIGURA 24.2 Histograma médio ampliado das freqüências versus tons G de amostras de 8 marcas cervejas, com resolução de 300 ppp no formato BPM. 55
- FIGURA 24.3 Histograma médio ampliado das freqüências versus tons B de amostras de 8 marcas cervejas, com resolução de 300 ppp no formato BPM. 56
- FIGURA 25 Gráfico dos *loadings* de PC1 e PC2 das marcas 8 de cervejas resolução de 600ppp, no formato BMP. As numerações referem-se aos valores dos índices de cor para os três canais justapostos. 58



## LISTA DE TABELAS

TABELA 1	- Algoritmo para extração de histogramas de freqüências dos índices de cor dos canais do sistema RGB.	61
TABELA 2	- Algoritmo de obtenção de histograma de freqüências.	62
TABELA 3	- Especificação das amostras das cervejas tipo pilsen digitalizadas nos formatos JPEG e BMP. A coluna “Local” representa a cidade de fabricação.	63
TABELA 4	- Especificação das marcas em resoluções 300 ppp e 600 ppp para o gráfico dos scores – Figura 15	64

**LISTA DE ABREVIATURAS E SIGLAS**

B	<i>Blue</i>
BMP	<i>Bitmap</i>
CCD	<i>Charge Coupled Device</i>
E.B.C	<i>European Brewertz Convention</i>
FDA	<i>Food and Drug Administration</i>
FAO	<i>Food and Agriculture Organization</i>
G	<i>Green</i>
JPEG	<i>Joint Picture Expert Group</i>
MIA	<i>Multivariate Image Analysis</i>
PCA	<i>Principal Component Analysis</i>
R	<i>Red</i>
SIP	<i>Scilab Image Processing</i>
RGB	<i>Padrão de Cores - Red, Green and Blue</i>

## LISTA DE SÍMBOLOS

2D	duas dimensões
3D	três dimensões
'	transposta da matriz
Cov	covariância
D	matriz diagonal
h	hora
i	linha
j	coluna
m	variáveis
mL	mililitro
n	amostras
P	matriz dos <i>loadings</i>
PC	Principal Component
ppp	pontos por polegadas
SVD	<i>Singular Value Decomposition</i>
t	autovalores
T	matriz dos <i>scores</i>
v	autovetores
X	matriz de dados originais
Xcm	matriz de dados originais centrados na média
W	matriz média de frequências nos índices RGB
VE%	percentual de variância acumulada pela componente principal

## RESUMO

Nesse trabalho foi desenvolvida uma nova metodologia para classificação de cervejas utilizando análise de imagens, tendo como atributo de qualidade a cor, com o emprego da Análise de Componentes Principais (PCA). Para o desenvolvimento do trabalho fez-se uso do *software* matemático *Scilab* e do pacote computacional *SIP*, que possibilitou a decomposição e a análise das imagens digitalizadas obtidas por meio de um *scanner* comercial. Rotinas foram aplicadas no *software* para decompor as imagens geradas nos formatos *JPEG* e *BMP*, em termos dos canais R, G e B, do padrão de cores RGB. Por meio dos algoritmos desenvolvidos, cada matriz de dados resultante da imagem de uma amostra de cerveja foi transformada em uma nova matriz, a qual gerou o respectivo histograma médio de frequências das 768 variáveis, cada canal de cor contém 256 (0 a 255) índices de cor. De tal modo, pôde-se analisar as diferentes marcas de cerveja por meio dos gráficos de PCA, com base nos resultados dos gráficos dos *loadings* e dos *scores* e dos histogramas médios de frequência de cada marca. Foram analisadas imagens em formato *JPEG* de dez marcas de cerveja do tipo *pilsen* em embalagens de latas contendo 350 mL de cerveja, escolhidas ao acaso, considerando as marcas de maior consumo na região metropolitana de Goiânia. Cerca de 200 mL de cerveja foram retiradas de cada lata e submetidas ao ultra-som para desgaseificação antes da captura da imagens no *scanner*. Durante as análises de imagens foram percebidas sensíveis alterações de cor das amostras de cerveja quando expostas a luz e ao ar, de modo que esse fenômeno passou a ser avaliado como sendo objetivo secundário do trabalho. Em um segundo momento, com o intuito de ampliar o estudo da análise de imagens, fez-se novas análises com imagens em formato *bitmap*, *BMP*, e para tanto foi realizada nova amostragem com o mesmo critério da anterior. Nessa nova amostragem foram escolhidas oito marcas, dentre as quais, sete coincidiram com as marcas utilizadas na análise das imagens com formato em *JPEG*. Os resultados obtidos nesse trabalho indicam que a metodologia proposta pode ser uma alternativa analítica simples para o controle de qualidade da cerveja, tendo como atributo a cor.

**Palavras-chaves: PCA, Análise de Imagens, Cerveja**

**ABSTRACT**

In this work a simple and new approach for lager beer pattern recognition based on image analysis was accomplished using Principal Component Analysis (PCA). Digital color images obtained from a scanner were decomposed into a binary form with the help of Scilab and SIP softwares. The PCA data matrix consisted of RGB color histograms for each sample. 350 mL beer cans were chosen at random on the shelf of local supermarkets in an overall of ninety cans from ten different brands. 200 mL samples were immersed in an ultrasound bath for a period of time ranging from twenty to thirty minutes before the digital image processing. The resulting methodology can be a simple alternative to the analytical quality control of lager beers.

**Keywords: PCA, image analysis, lager beer**

## 1-INTRODUÇÃO

A cerveja é uma bebida de ampla difusão e intenso consumo, sendo conhecida desde remota antigüidade em diversos países do mundo, principalmente nos continentes Europeu e Asiático. Tem como principal mercado consumidor a Europa, sendo que a Alemanha, República Tcheca e a Dinamarca possuem o maior consumo *per capto* em litros por ano. Historicamente, há 2800 a.C., os babilônios já fabricavam vários tipos de cervejas. Diversos documentos antigos relatam histórias e lendas ligadas à cerveja dos povos hititas, armênios, gregos, egípcios e outros. Em várias regiões a bebida era utilizada na alimentação diária da população como importante fonte de nutrientes. (CERVESIA, 2006).

Em nosso país a cerveja chegou por volta de 1808, trazida por Dom João VI. Até meados do século XIX ela era importada sendo considerado privilégio dos nobres. Na atualidade o Brasil é o quarto produtor mundial, com um consumo *per capto* anual de 50 Litros. (MATHIAS, 1997)

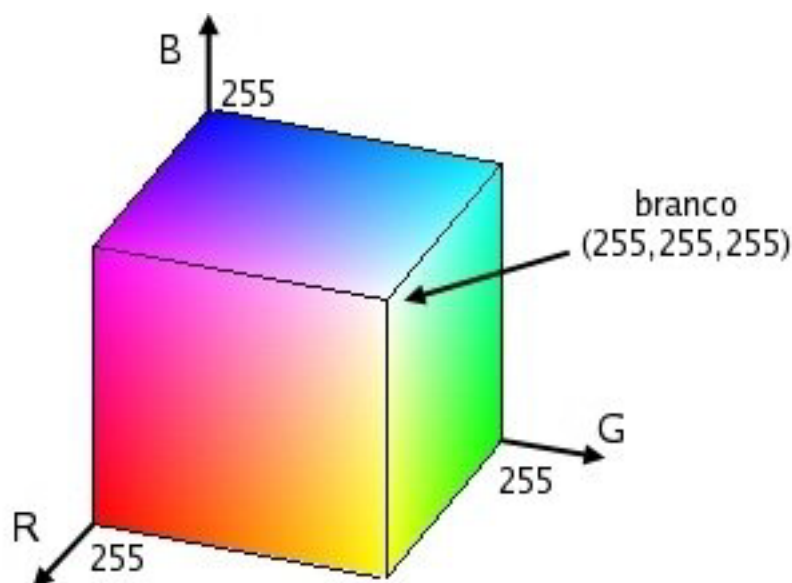
A cerveja é obtida através da ação fermentativa da levedura *cervejeira* (fermento) sobre o mosto, que é constituído pela mistura de água, malte e lúpulo. Comercialmente a cerveja mais encontrada no mercado é a do tipo *pilsen* (de baixo teor alcoólico), com cor clara, levemente amarelada, com poucos carboidratos fermentáveis, originalmente fabricada com malte de cevada e utiliza água com baixo teor de sais dissolvidos. A cerveja é uma bebida antiga, cujo processamento sofreu diversos melhoramentos em cada etapa de produção, possibilitando o domínio da tecnologia *cervejeira* garantindo também uniformidade e qualidade na produção, características fundamentais para a conquista do mercado consumidor. Apesar do Brasil não ser o maior consumidor da bebida, o país tem um grande destaque no mercado e isso serve de incentivo para a melhoria e o progresso contínuo das técnicas utilizadas na indústria *cervejeira*. Legalmente no Brasil, o mercado de bebidas é regido pelas seguintes legislações: LEI nº 8918/94

regulamentada pelo DECRETO nº 2314/97 que especifica das matérias-primas utilizadas na fabricação, incluindo a composição final da cerveja. Nesse decreto constam a definição e a classificação da cerveja quanto às especificações do tipo de fabricação, da cor e do teor alcoólico, dos padrões de qualidade, padrões de rotulagem e análises de controle. (CERVESIA, 2006).

Como forma de melhoria das técnicas de fabricação cervejeira, a implementação de métodos novos, acessíveis e rápidos para o controle de qualidade e a criação de um padrão de identidade é de grande importância para a indústria de cerveja. Dentre os atributos da bebida, um dos que facilmente pode ser caracterizado, é a cor da cerveja tipo *pilsen*. Até o presente momento a classificação de cor em cerveja é realizada durante o processo de produção, por meio de uma escala padrão Européia conhecida como E.B.C. (*European Brewerz Convention*). Segundo essa escala, a cerveja clara deve conter menos de 20 unidades E.B.C. e igual ou superior a 20 unidades para cerveja escura. Sabe-se que essa escala não é claramente definida, e para que seja realizada a medição devem ser adquiridos equipamentos específicos.

A cor amarelada da cerveja é devido aos pigmentos conhecidos como melanoidinas e aos caramelos presentes no malte, além dos adjuntos usados durante o tratamento térmico do mosto devido à reação de Maillard, um tipo de oxidação não enzimática do processo de caramelização.

Como forma alternativa proposta nesse trabalho, a classificação da cerveja por meio da cor utiliza a escala de padrão de cores RGB (*red, green e blue*) que utiliza 256 tons de cores (índices de cor), que variam entre 0 e 255 tons de cores para cada uma das cores básicas: vermelha(R), verde(G) e azul(B), totalizando 768 tons que podem ser tratados como variáveis, permitindo uma combinação de  $256 \times 256 \times 256$ , alcançando um total de 16.777.216 tonalidades de cada *pixel* (elemento de imagem digital). Nesse formato, um tom de cor corresponde a um ponto em um espaço tridimensional formado pelos eixos R, G e B (Figura 1). Essa escala é empregada em imagens digitais capturadas por meio de *scanners* ou máquinas fotográficas digitais (SCHIMIDT, 1997).



**Figura 1** – Cubo de tons dos canais RGB

Imagens digitais têm sido usadas numa ampla faixa de aplicações, como por exemplo: avaliação de defeito em semicondutores, medida de distribuição de ingrediente ativo em comprimidos farmacêuticos, detecção de traços de contaminantes em células de combustível, imagens de satélites, imagens médicas (clínicas) e etc. (YU e MACGREGOR, 2003)

Segundo a definição de GELADI e GRAHN (1996), imagem é uma reprodução de um objeto real ou cena, preservada em um meio. Para propósitos científicos, as imagens sempre são geradas com o objetivo de expressar algumas propriedades objetivas do objeto ou cena de interesse. Para que uma imagem seja utilizada em trabalhos científicos, esta deve ser expressa como uma função matemática. No entanto essa função matemática contínua que descreve a imagem deve ser convertida em uma função discreta através da digitalização da imagem. No processo de digitalização uma imagem contínua é transformada em uma imagem digital que consiste em uma estrutura quadriculada, onde cada quadrado recebe o nome de *pixel*, o qual possui um valor de intensidade correspondente. A resolução da intensidade de um *pixel* numa imagem digital pode ter vários valores dependendo da área de aplicação. Historicamente, tem-se utilizado



valores inteiros, já para imagens químicas são utilizados valores reais com precisão decimal de até oito dígitos. Deve-se ter em mente que quanto maior a resolução espacial e de intensidade de uma imagem digital, maiores serão os recursos computacionais necessários para seu processamento.

A cor é uma propriedade de grande importância nas indústrias alimentícias, têxtil, de fotografia, em publicações, na propaganda, e em outras aplicações. As cores podem ser simuladas através da mistura das três cores primárias. Na análise de imagens, geralmente, é realizada a decomposição da imagem colorida em imagens de cores primárias gerando três imagens que serão processadas como uma imagem multivariada. Várias operações podem ser realizadas durante o processo de análise de imagens multivariadas. Se as operações levam à redução dos dados para a de forma que se permita chegar a conclusões sobre o sistema, dá-se o nome de análise de imagens.

Um processamento normalmente utilizado nas imagens multivariadas é o escalamento linear ou então o escalamento não linear. O escalamento é, normalmente, conhecido como uma operação de pré-tratamento dos dados. Exemplos de escalamento linear são: centrar na média, dividir pelo desvio padrão e pela variância. O escalamento não linear geralmente é aplicado quando se tem um modelo físico que apresenta uma resposta não linear, como por exemplo, uma imagem que apresenta um histograma assimétrico. Uma das ferramentas mais importantes no estudo de imagens multivariadas é a análise de componentes principais (PCA). O cálculo da PCA de uma matriz de imagens é realizado através dos autovalores e autovetores da matriz (LAQQA, 2007)

Imagens digitais podem ser convertidas em matrizes de dados por meio de aplicativos matriciais, como o Scilab (SIP, 2005), com *toolbox* SIP (*Scilab Image Processing*) (SIP, 2003). As matrizes de dados geradas pelas imagens digitais podem também ser decompostas em histogramas dos canais R, G e B, específicos para cada amostra. A partir da matriz de dados das imagens é obtido a PCA, método estatístico multivariado, caracterizado pela redução de um espaço multidimensional, através da combinação linear das variáveis originais. Em outras palavras, a PCA é uma metodologia

exploratória que visa evidenciar similaridades ou diferenças entre as amostras, comparando os seus conjuntos de dados (MINGOTI, 2005).

O uso de imagens digitais tem se difundido, pois está associado ao desenvolvimento de diversos setores como: aplicações industriais, comunicações e pesquisas (PRATT, 1991). Em Química o uso de imagens digitais é pequeno, mas suas aplicações vêm crescendo nos últimos anos, podendo ser citados como exemplos de usos: na determinação fotométrica de compostos em matrizes de papéis (SCHIMIDT, 1997) e por cromatografia em camada delgada (HAYAKAWA & HIRAI, 2003), na determinação do teor de saponina em quinoa (SOUZA *et al*, 2004), no estudo do fluxo de L-Glutamato em cérebro de ratos (HIRANO *et al*, 2003), cromatografia gasosa bi-dimensional (REICHENBACH *et al*, 2004), alimentos, entre outras (APEXBRASIL, 2005).

## **1.1- OBJETIVO**

O presente trabalho tem por finalidade propor metodologia alternativa, acessível e de baixo custo para a classificação de cervejas do tipo *pilsen*, que tem como base a cor (atributo de qualidade), com o emprego da PCA dos histogramas médios de freqüências dos índices de cor das imagens das cervejas obtidas com um *scanner* comum que possibilita a comparação de diferentes marcas de cervejas com base nas similaridades de cada marca.

## 2-FUNDAMENTOS TEÓRICOS

### 2.1- CERVEJA

A cerveja é uma bebida de ampla difusão e intenso consumo, sendo conhecida desde a antigüidade em diversos países do mundo. A fabricação de cerveja passou de uma atividade artesanal e converteu-se em um ramo economicamente muito importante dentro da microbiologia industrial. Depois da legislação sobre pureza, que entrou em vigor em 1516 (Baviera - Alemanha), a cerveja só pode ser fabricado a partir de malte, cereais, lúpulo, água pura e leveduras. (CERVESIA, 2006) Como as leveduras da cerveja não transformam o amido em açúcar, esse não pode ser levado direto à fermentação. Faz-se necessário, primeiro transformar o amido dos grãos do cereal, submetendo-os à maceração e à germinação, sendo tostados em malte rico em açúcar, a partir do qual se irá obter, por maceração e cocção em água, o mosto de cerveja estéril. Esse fermentará gerando cerveja somente após as etapas de adição de lúpulo, resfriamento e inoculação com as leveduras da cerveja.

Os aspectos utilizados para a classificação, assim como as características sensoriais de uma cerveja, dependem basicamente da sua natureza e das características das matérias-primas utilizadas, além do tipo de levedura e da condução do processo de fermentação, que terão influência direta no tipo de cerveja obtida. (CERVESIA, 2006).

Os tipos de cerveja estão relacionados com:

- Conteúdo alcoólico;
- Tipo de fermentação;
- Concentração do malte e do lúpulo utilizado;
- Duração do envelhecimento;
- Teor do extrato final;
- Extrato primitivo.

As cervejas são classificadas segundo algumas características do produto e o tipo de processo utilizado na sua fabricação. Os principais aspectos levados em consideração para a classificação da cerveja são:

#### A - Quanto ao Tipo de Fermentação:

- Alta Fermentação (12-15°C)

São as cervejas antigas, isto é, as produzidas antes do domínio tecnológico da fermentação. Sua fabricação leva a concentrações elevadas de malte e lúpulo e a um envelhecimento de maior duração. Esse tipo de cerveja é obtido pela ação da levedura cervejeira, que surge à superfície da fermentação tumultuosa (flotante) devido à retenção de gás pelas leveduras; a coleta do fermento é feita nessa etapa do processo.

Esse tipo de fermentação fornece cervejas de maior teor alcoólico, podendo citar-se *Weizenbier*, que é produzida com trigo maltado, normalmente servido com a levedura em suspensão, sem filtragem, e *Stout* que é uma cerveja forte, escura e rica em lúpulo, possui alta graduação alcoólica chegando a 5,6%.

- Baixa Fermentação (5-10°C)

É a cerveja obtida pela ação da levedura cervejeira, que se deposita no fundo do tanque (floculante), após a fermentação tumultuosa. Pode ser citada como de baixa fermentação a cerveja tipo *Pilsen*, com cor clara, poucos carboidratos fermentáveis, fabricada com 100% de malte de cevada e utiliza água com baixo teor de sais dissolvidos.

#### B - Quanto ao Teor de Extrato Primitivo:

O teor de extrato primitivo é a densidade original do mosto, antes da fermentação.

#### C - Quanto a Cor

A cerveja também pode ser classificada quanto à cor. Essa classificação é através da escala E.B.C. (*European Brewers Convention*). A cerveja clara deve conter menos de 20 unidades E.B.C. e a cerveja escura quantidade igual ou superior a 20 unidades E.B.C.

#### D - Quanto ao Teor Alcoólico:

Essa classificação é realizada através da medida da porcentagem de álcool existente na cerveja. As classes são:

- Cerveja sem Álcool: menor ou igual a 0,5%;
- Cerveja de Baixo Teor Alcoólico: maior que 0,5% e menor que 2,0%;
- Cerveja de Médio Teor Alcoólico: igual ou maior que 2,0% e menor que 4,5%;
- Cerveja de Alto Teor Alcoólico: igual ou maior que 4,5% e menor que 7,0%.

Teoricamente, uma cerveja classificada como: "sem álcool" não deveria conter nenhuma quantidade de álcool etílico, mas na realidade, há um pequeno teor de álcool remanescente que diverge de acordo com as legislações próprias dos vários países produtores. A legislação internacional orienta no sentido de admitir um teor de álcool máximo de 0,5% em volume, esse tipo de cerveja. A qualidade da cerveja sem álcool evoluiu muito quando novas tecnologias diminuíram as características negativas no paladar, inerentes aos processos originais de fabricação, com fermentação interrompida. Atualmente os processos mais utilizados são os que iniciam de cervejas normais separando álcool, seja através de sua evaporação sobre vácuo, seja por meio do uso de membranas especiais.

#### E - Quanto ao Teor de Extrato Final:

É o percentual de extrato que foi fermentado na fabricação de determinado tipo de cerveja.

As classes são:

- Baixo teor de extrato: teor de extrato menor que 2%;
- Médio teor de extrato: teor maior ou igual a 2% e menor que 7%;
- Alto teor de extrato: teor de extrato maior ou igual a 7%;

#### Tipos de Cerveja

Nos dias atuais, estima-se, que existam cerca de 20 mil marcas e 150 tipos de cerveja, no mundo. Nacionais ou importadas se dividem em três grandes famílias: *Lager*, *Ale*, e *Wheat*.

Qualquer alteração nos ingredientes ou no processo implica em uma enorme variedade de sabores, sendo possível escolher a cerveja pelo paladar: forte, fraca, clara, escura, amarga, doce.

### *LAGER*

As cervejas da família das *Lagers* são leves, de diversas graduações alcoólicas. É a mais comumente encontrada no Brasil, sendo que todas as fábricas grandes ou pequenas produzem cervejas tipo *Lager*.

A maior parte da produção é de cerveja do tipo *pilsen*; seguida de *malzbier*, *münchen* e *lager* brasileira, além da bock, cuja produção cresceu muito na década de 1990 com posterior declínio na década seguinte.

As Cervejas tipos *Lager* são produzidas com leveduras de baixa fermentação. Podem ser:

- *Pilsen* ou *Pilsner*: Originária da cidade de Pilsen na Boêmia, República Tcheca, cerveja encorpada e de sabor forte. No Brasil e em outros países é mais leve, devido à utilização de milho e arroz. Tem de 3 a 5 graus alcoólicos (leve), e deve ser tomada entre 4 e 6 °C. Todas as marcas brasileiras produzem pelo menos uma variedade de cerveja do tipo *pilsen*.

- *Dortmunder*: Variação da *pilsen* produzida na Alemanha. É uma cerveja clara e menos amarga que a *pilsen* original, utiliza água e um teor elevado de sulfato de cálcio, extrato primitivo 12 a 14%.

- *München*: pode ser clara ou escura e com leve paladar picante de malte e café, com 4 a 5 graus de álcool. Originária da cidade de Munique na Alemanha.

- *Malzbier*: cerveja escura e doce, que pode ser produzida com malte mais tostado ou com caramelo em sua fórmula.

- Bock: cerveja forte com 5 a 7 graus alcoólicos de cor avermelhada. Contém maltes especiais, torrados e caramelados, sabor pouco amargo. Há ainda as cervejas doppelbock, da Alemanha, consideradas mais fortes (13,2 graus alcoólicos) reconhecidas pelos nomes com sufixo "ator" (*Delicator*, *Celebrator*, *Kulminator*).

- Draft: com 4 a 4,5 graus alcoólicos, é uma cerveja típica americana. Devido a um processo especial de filtragem, consegue-se uma

cerveja mais leve, menos alcoólica e amarga. O tipo de lúpulo confere resistência à luz, às draft podem ser envasadas em garrafas transparentes.

- *Lager Americana*: versão da *pilsner*, muito leve, quase sem lúpulo e praticamente com pouco sabor. Denominada também, *Standard Lager Americana*.

- *Lager Brasileira*: semelhante à *Lager Americana*, com mais lúpulo e adjuntos como milho e o arroz, muito próximo das cervejas escandinavas.

### *WHEAT*

As cervejas *wheats* podem ser claras ou escuras. Elaboradas com milho, trigo e até frutas. Resultam em bebidas ou muito leves ou muito pesadas. O teor de álcool varia de 2,5 a 7 graus.

### *ALE*

As *Ale* são cervejas aromáticas em geral escuras e com forte paladar amargo devido ao lúpulo. Totalizam 5% da produção mundial, que se concentra na Alemanha, Inglaterra e Bélgica. Teor alcoólico em volume de 3 a 4%. As cervejas do tipo *Ale* são produzidas com leveduras de alta fermentação.

## Matérias Primas da Cerveja

### Água

A água é a matéria-prima mais importante na fabricação de cerveja, de modo que suas características físico-químicas são de fundamental importância para se obter uma cerveja de boa qualidade.

Dependendo da origem da água, todas ou apenas algumas das seguintes operações são efetuadas na cervejaria:

- Aeração - oxidação para remover odores;
- Clarificação: adição de produtos químicos para a aglomeração ou coagulação de material em suspensão, que será decantado ou filtrado;
- Filtração: remoção de sólidos em suspensão, filtrando-se sobre a areia;
- Cloração: para eliminação de microorganismos;

- Desmineralização: para a remoção de sais em águas que contenham alto teor de sais dissolvidos.

### Malte

O malte é um produto rico em açúcar, obtido com a germinação parcial dos grãos de cereais. A princípio, qualquer cereal pode ser maltado, como por exemplo: o milho, trigo, centeio, aveia e a cevada. A escolha, entretanto, deve levar em consideração, entre outros fatores, o poder diastásico e o valor econômico de cada cereal. O cereal mais utilizado para a produção de cerveja é a cevada, apesar de diversos outros cereais poderem ser utilizados também.

O processo de maltagem consiste na germinação controlada de grãos, e para a avaliação do potencial de maltagem, devem ser analisadas algumas características do grão, dentre os quais: complementos do malte.

Os complementos do malte incluem os açúcares menores, amido, cereais, raízes e tubérculos. Estes complementos são utilizados pelos seguintes motivos: produção de cervejas mais leves que as produzidas com malte puro; melhoramento da estabilidade coloidal, evitando a floculação e turvação; e diminuição de custos.

A legislação dos diferentes países europeus, quanto ao emprego de adjuntos, pode ser dividida em três grupos. No Brasil, a legislação não limita o percentual de arroz, *gritz* de milho e xarope de milho. No entanto, essa legislação limita o emprego de açúcar em cervejas claras, que deve ficar entre 10 e 15% em relação ao extrato primitivo e 50% no caso de cervejas escuras. (BRASIL 1, 1994)

### Lúpulo

O lúpulo é o responsável pelo aroma acre e sabor amargo característicos da cerveja. Além de evitar a formação de espuma durante a fervura, ainda atua como agente bacteriostático.



## Adjuntos de Fabricação

### Antioxidantes

Têm como função evitar a ação do oxigênio, que é o principal fator da deterioração das gorduras dos alimentos. Quando sofrem transformações, essas gorduras acabam por alterar o sabor e odor dos alimentos, tornando-os impróprios para o consumo.

### Estabilizantes

Mantém as características físicas das emulsões e suspensões. São adicionadas as cervejas com a finalidade de aumentar sua viscosidade.

### Acidulantes

Atuam como adjunto de aroma e sabor, encontrados na natureza, disponíveis comercialmente na forma de soluções aquosas incolores, inodoras, viscosas e não voláteis. São geralmente produzidos pela fermentação de sacarose altamente refinada. O ácido láctico é o acidulante amplamente utilizado na indústria cervejeira. Este componente auxilia no ajuste do pH, obtendo rendimento máximo no extrato e diminuindo o tempo de maceração, favorecendo a precipitação das proteínas e melhorando a cor e a filtração do mosto. O uso do ácido láctico é recomendado para alimentos e bebidas por instituições internacionais como a FDA norte americana e o FAO/WHO da Organização das Nações Unidas.

### Antiespumantes

Atuam regulando a formação de espumas, principalmente durante a fervura.

### Leveduras

As leveduras mais utilizadas em cervejaria são de duas espécies do gênero *Saccharomyces*: *Saccharomyces cerevisiae* (alta fermentação) e *Saccharomyces uvarum* (baixa fermentação). Uma levedura de baixa fermentação é considerada de boa qualidade para a produção de cerveja, se permanecer em suspensão durante a fase ativa da fermentação e então

flocular e sedimentar, favorecendo a separação rápida da cerveja clarificada do sedimento.

### Processamento da Cerveja

O processo de fabricação de cerveja é dividido nas seguintes etapas:

Moagem do Malte; Mosturação, Brassagem, Decocção, Filtração do Mosto, Fervura do Mosto, Tratamento e Resfriamento do Mosto, Preparo do Inoculo, Fermentação, Maturação, Acabamento (Clarificação e Carbonatação) e Embalagem.

## 2.2-IMAGENS DIGITAIS

Imagem é a reprodução de um objeto real ou cena, preservada em um meio. Podem-se citar como exemplos de imagens pintura a óleo, aquarela, fotografia ou escultura. Para propósitos científicos, as imagens sempre são geradas com objetivo de expressar algumas propriedades objetivas do objeto ou cena de interesse. As imagens são utilizadas na investigação científica há muito tempo, inicialmente apenas como ferramentas para a representação da prática científica. (LYNCH & WOOLGAR, 1990). Para que seja utilizada como objeto de estudo da ciência, uma imagem deve ser expressa como uma função matemática. No entanto, essa função matemática é muito complexa para uma imagem contínua, o que torna sua obtenção e tratamento tarefas de difícil execução. Desse modo, a função matemática contínua que descreve a imagem é convertida em uma função discreta através da digitalização da imagem. No processo de digitalização uma imagem é transformada em uma imagem digital, que consiste em uma estrutura quadriculada onde cada quadrado recebe o nome de *pixel*, o qual possui valores de intensidade correspondente. Os *pixels* podem ter perfil quadrado, retangular ou qualquer outro formato, mas no geral são utilizados *pixels* quadrados. As imagens digitais podem ter dimensões variadas, tais como duas (2D), três (3D) ou mais (4D). Como exemplos de imagens 3D podem-se citar as imagens de ressonância magnética nuclear. O número de *pixels* que forma a imagem digital define a sua resolução espacial. A

resolução da intensidade de um *pixel* numa imagem digital pode ter vários valores dependendo da área de aplicação. Em geral, para imagens químicas são utilizados valores reais com precisão decimal de até 8 dígitos e para imagens fotográficas são utilizados valores inteiros de 0 a 255. Deve-se ter em mente que quanto maiores forem às resoluções espaciais e de intensidade de uma imagem digital, maiores serão os recursos computacionais necessários para seu processamento. (LAQQA, 2007).

Quando uma imagem é capturada por algum dispositivo e transferida para um computador, ela é inicialmente decodificada em pontos e cada um desses pontos é devidamente armazenado em uma posição de memória. Posteriormente, estes pontos são transferidos para o monitor de vídeo do computador segundo critérios pré-estabelecidos, de modo a reconstruir a imagem.

Os dispositivos de aquisição de imagens mais comumente empregados em conjunto com os computadores são as câmeras fotográficas digitais e os *scanners*. Esses equipamentos utilizam detectores eletrônicos baseados em dispositivos acoplados por carga (CCD), formados por arranjos de semicondutores distribuídos em uma seqüência linear ou em uma matriz retangular.

Uma imagem a ser digitalizada é decomposta em um conjunto de pontos, os quais podem ser tratados individualmente. Em geral, o conjunto de pontos é sempre tratado na memória do computador como uma matriz de  $n$  linhas e  $m$  colunas e seu tamanho depende de alguns fatores, como por exemplo, o número de elementos que compõem o detector de CCD. Outro importante fator é a natureza da imagem, que pode ser monocromática ou policromática. Nas imagens coloridas utilizadas em monitores de TV e computador, as cores são, normalmente, representadas pela combinação de três imagens univariadas. Uma para o vermelho, uma para o verde e uma para o azul, cores que formam padrão RGB. Cada pixel é representado por três bytes de oito bits cada, totalizando 24 bits. Como um byte pode expressar inteiros entre 0 e 255, teremos um total de  $256 \times 256 \times 256$ , ou seja, 16 777 216 cores possíveis (ou  $2^8 \times 2^8 \times 2^8$ ). (GELADI & GRAHAN, 1996).

Os *scanners* utilizam um único conjunto linear de detectores tipo CCD, de modo que para a aquisição da imagem é necessário que os seus detectores sejam deslocados sobre o objeto. Equipamentos de aquisição de imagens por meio de varredura, como os *scanners* possuem uma fonte de luz própria e a radiação emitida por esta fonte, após ser refletida pelo objeto, é direcionada por um conjunto de lentes para um arranjo de detectores. A cada passo de deslocamento do *scanner* sobre o objeto, os dados da imagem obtidos pelo conjunto de detectores são transferidos para uma interface na forma de sinais analógicos. Estes dados são então convertidos em sinais digitais, resultando em um fragmento do objeto, representando uma linha da matriz de dados de sua imagem. Esse processo se repete até que toda a imagem do objeto seja adquirida, resultando na digitalização total da imagem, representada por uma matriz  $X$  com  $i$  linhas e  $j$  colunas. Cada coluna desta matriz foi obtida por um mesmo detector que compõem o arranjo de CCD e cada linha representa a posição do conjunto de detectores no momento de aquisição. Cada elemento  $C(i,j)$  dessa matriz corresponde a um elemento da figura ou *pixel*. Os índices  $i$  e  $j$  são utilizados para determinar os endereços de cada *pixel* da imagem do objeto durante a etapa de sua construção na tela do monitor de vídeo.

Os *scanners* podem digitalizar uma imagem com diferentes resoluções. No processo de aquisição, quanto maior for a resolução, maior será o número de pontos digitalizados por área, conseqüentemente haverá uma melhor definição da imagem, resultando em um aumento da matriz de dados que representa a imagem. Outro fator que determina o tamanho da matriz no momento da aquisição da imagem é o formato de cor escolhido, o qual determina o número de bits por *pixel* da matriz. (SCHIMIDT, 1997).

A cor é uma propriedade de grande importância nas indústrias alimentícias, têxtil, fotográfica, em publicações, na propaganda, e muitas outras aplicações. As cores podem ser simuladas através da mistura das três cores primárias. Conforme descrito anteriormente, para o caso dos monitores de computador e de televisores são utilizadas as três cores, padrão RGB. Nos processos de impressão é utilizado um padrão diferente de cores, o ciano, o carmim e o amarelo.

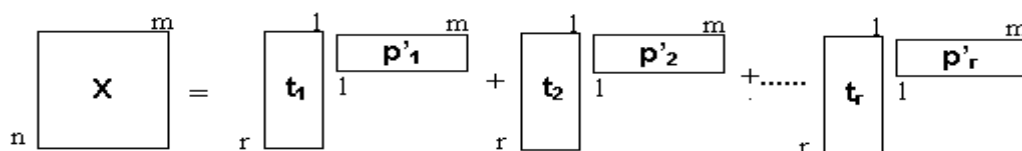
## 2.3-ANÁLISE DAS COMPONENTES PRINCIPAIS (PCA)

A análise de componentes principais PCA (*Principal Component Analysis*), é um método para decompor uma matriz de dados  $\mathbf{X}$  de posto  $r$  (ou ranque  $r$ ), como uma soma de matrizes de posto igual a 1, onde posto é um número que expressa a dimensão de uma matriz.

Essas novas matrizes de posto 1, são produtos de vetores chamados “scores”  $t_h$  e “loadings”  $p_h$ . Esses “scores” e “loadings” podem ser calculados par a par, em um processo iterativo, como na Equação 1.

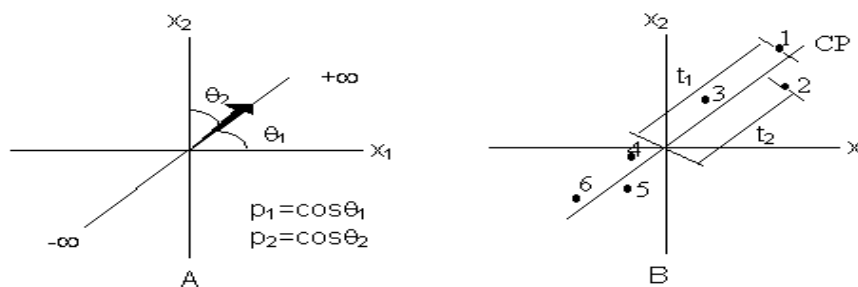
$$\mathbf{X} = t_1 p_1' + t_2 p_2' + \dots + t_r p_r' \quad (1)$$

A Figura 2 traz a representação da matriz de dados decomposta em matrizes “scores” e “loadings”.



**Figura 2** – Representação da matriz de dados  $\mathbf{X}$  decomposta em produto de matrizes de posto igual a 1.

Para exemplificar  $t_h$  e  $p_h'$ , temos a Figura 3 que ilustra nos planos bidimensionais duas variáveis  $x_1$  e  $x_2$ . A Figura 3 (A) mostra uma componente principal que é a reta que aponta para a direção de maior variabilidade das amostras da Figura 3 (B). Os “scores”  $t_h$  são as projeções das amostras na direção da componente principal e os  $p_h'$  “loadings” são os cossenos dos ângulos formados entre a componente principal e cada variável. (LAQQA, 2007)



**Figura 3** - Uma componente principal de duas variáveis: (A) *loadings*: ângulos do vetor direção; (B) *scores*: projeções das amostras (1 a 6) na direção da componente principal. Dados são centrados na média.

A análise de componentes principais é uma das bases fundamentais da maioria dos métodos para tratamento de dados multivariados. A PCA consiste em uma mudança de base, para melhor descrever as informações relevantes em uma calibração multivariada, utilizando autovetores que oferecem uma interpretação matemática mais simplificada.

O problema de se trabalhar com muitas variáveis para resolução de problemas em química deve-se à complexidade dos dados e à dificuldade de se extrair informações relevantes de dados isolados, uma vez que sem uma correlação das variáveis medidas não é possível avaliar o problema em uma visão macro. Se, de um lado, o aumento do número de informações pode representar um maior conhecimento do sistema em estudo, por outro, pode ocorrer à inserção de um grande número de informações irrelevantes ou redundantes.

Quanto maior o número de dados de um sistema de análise química, maior será o número de dimensões requeridas para representar esse sistema (NOBLE e DANIEL, 1986). Se os dados estão inseridos em um espaço vetorial com grande dimensão, a interpretação dos efeitos das variáveis torna-se problemática, devido ao grande número de variáveis que devem ser monitoradas. Entretanto, para se obter a interpretação de um sistema complexo, devem-se procurar subsistemas menores que representem e simplifiquem o sistema complexo em estudo. Isso, na verdade, significa encontrar um vetor,  $\mathbf{v}$ , chamado autovetor, e um escalar,  $t$ , chamado de autovalor, que satisfaça a igualdade  $\text{cov}(\mathbf{X})\mathbf{v} = t\mathbf{v}$ . Dessa equação encontra-se o auto-sistema da covariância dos dados ( $\text{cov}(\mathbf{X})$ ), de maneira que transforme um espaço polidimensional em um subespaço de dimensão um gerado por um autovetor, (NOBLE e DANIEL, 1986). ou seja, uma matriz,  $\mathbf{X}$  (multivariável), representada por muitos vetores, passa a ter as informações relevantes, representada por um único autovetor. Esse subsistema retém as informações relevantes e representam mais adequadamente cada variável individualmente, de modo a facilitar a interpretação do sistema. (MESSERSCHIMIDT, 1999).

A PCA utiliza-se da diagonalização da matriz de dados, ou seja, do estudo dos autovalores associados aos autovetores para filtrar informações relevantes do sistema em estudo. A matriz diagonal da matriz de dados é um operador linear mais simples possível que representa uma matriz de dados. Efetuando-se a diagonalização da matriz de dados, encontram-se os autovetores que explicarão a maior variância desses dados associados aos maiores autovalores. Os autovetores podem ser entendidos como sendo o peso da transformação de uma mudança de base de um sistema e os autovalores as raízes do problema desse sistema. NOBLE e DANIEL (1986) relatam um estudo de auto-sistemas (autovetores e autovalores) aplicado ao movimento harmônico simples (vibração), onde os autovalores representavam tipos distintos de vibração, e os autovetores as modas dessas vibrações.

Em analogia com a mudança de base, as coordenadas antigas podem ser representadas por uma soma linear do produto de novas coordenadas (*scores*),  $\mathbf{t}$ , e pesos da transformação (*loadings*),  $\mathbf{p}$ , de modo que, a variância capturada pela primeira componente principal, **PC1**, seja tão grande quanto possível como descrita na Equação 1. (GELADI e KOWALSKI, 1986)

$$PC1 = t_1 p_1^T + t_2 p_2^T + \dots + t_n p_n^T + \sum_{i=1}^n e_i \quad (1)$$

Devido a essa soma linear, cada amostra será representada por um único valor, substituindo as multivariáveis originais. Por esse motivo a PCA tem capacidade de reduzir a dimensionalidade dos dados (MELLO, 1998), o que é importante nos sistemas em que:

- a) faltam seletividade de variáveis;
- b) as variáveis são redundantes;
- c) existe falta de conhecimento químico do sistema analisado.

No entanto, se a primeira componente principal, **PC1**, não explicar o comportamento da variância dos efeitos causados no sistema, uma segunda componente principal, **PC2**, ortogonal à primeira deve ser utilizada para completar a explicação da variância. À medida que se utilizam mais componentes principais, sem um estudo do efeito causado no sistema de previsão e concentração de amostras desconhecidas, ruídos e informações irrelevantes podem ser adicionados no modelo de calibração (FERREIRA *et al.*, 1999).

A representação matricial da equação 2 pode ser feita pela equação abaixo:

$$\mathbf{X}=\mathbf{TP}^T \quad (2)$$

Outra maneira de se obter os autovetores (*loadings*) e autovalores (*scores*) ocorre por um tipo especial de diagonalização, a Decomposição em Valores Singulares, DVS (**SVD**). A **DVS** possibilita representar facilmente uma matriz **B** (dados) complexa usando novas bases ortonormais, uma para o espaço domínio (**v**) e outra para o espaço imagem (**u**) de **B** (NOBLE e DANIEL, 1986). Essa transformação pode ser definida por meio da matriz **B** por **B(w)= B.w**. Mudando-se de base no espaço domínio de **B** relativamente aos novos vetores básicos, **v**(V1, ....Vn), de **w** para **w'**, tem-se:

$$\mathbf{w}=\mathbf{v}.\mathbf{w}' \quad (3)$$

$$\mathbf{w}'=\mathbf{v}^T.\mathbf{w}$$

Mudando de base no espaço imagem de **B** relativamente aos novos valores básicos, **u** (u<sub>1</sub>, ..., u<sub>n</sub>), de **q** para **q'**, tem-se a (4)

$$\mathbf{q}=\mathbf{u}.\mathbf{q}'$$

$$\mathbf{q}'=\mathbf{u}^T.\mathbf{q} \quad (4)$$

Fazendo **q=B.w** e aplicando-se a equação anterior tem-se (5)

$$\mathbf{q}=\mathbf{B}.\mathbf{v}.\mathbf{w}' \quad (5)$$

Fazendo a substituição da equação 4, utilizando-se a equação 5, obtém-se (6)

$$\mathbf{q}'=\mathbf{u}^T.\mathbf{B}.\mathbf{v}.\mathbf{w}' \quad (6)$$



O operador linear, **S**, é a matriz mais simples que representa **B** usando novas bases ortogonais (vetores **v** e **u**), de acordo com (7)

$$\mathbf{S} = \mathbf{u}^T \cdot \mathbf{B} \cdot \mathbf{v} \quad (7)$$

A equação acima é uma equação de semelhança, pois **S** é semelhante a **B**, ou seja, **S** é a matriz que contém as informações sobre a quantidade de variância que cada componente principal descreve. Isolando **B** da equação acima, o mesmo é decomposto e representado por (8)

$$\begin{aligned} \mathbf{u} \cdot \mathbf{S} \cdot \mathbf{v}^T &= \mathbf{v} \cdot \mathbf{v}^T \cdot \mathbf{B} \cdot \mathbf{v} \cdot \mathbf{v}^T \\ \mathbf{u} \cdot \mathbf{S} \cdot \mathbf{v}^T &= \mathbf{B} \quad (8) \end{aligned}$$

onde o produto de **u.S** representa a matriz de *scores*, **t**, e **v** representa a matriz de *loadings*. Aplicando cálculo de variância-covariância, **B<sup>T</sup>B**, tem-se (9):

$$\begin{aligned} \mathbf{B}^T \mathbf{B} &= (\mathbf{U} \cdot \mathbf{s} \cdot \mathbf{v}^T)^T \cdot (\mathbf{u} \cdot \mathbf{S} \cdot \mathbf{v}^T) \\ \mathbf{B}^T \mathbf{B} &= \mathbf{v} \cdot \mathbf{S}^T \cdot \mathbf{u}^T \cdot \mathbf{u} \cdot \mathbf{S} \cdot \mathbf{v}^T \\ \mathbf{B}^T \mathbf{B} &= \mathbf{v} \cdot \mathbf{S}^T \cdot \mathbf{S} \cdot \mathbf{v}^T \quad (9) \end{aligned}$$

Verifica-se que na equação acima a matriz **v** fornece os autovetores associados aos autovalores de **S<sup>T</sup>S**. Nesse caso, s elementos da diagonal da matriz obtida pelo produto das matrizes **S<sup>T</sup>S**, contém as informações sobre a quantidade da variância ( $s_{ij}^2$ ) dos dados espectrais que a i-ésima componente principal descreve. Da mesma forma, **BB<sup>T</sup>** resulta em **SS<sup>T</sup>** que contém as informações sobre a variância e covariância associada aos dados do vetor **u**. Para determinar a porcentagem da explicação da variância, VE(%), pelas componentes principais utiliza-se a Equação nº 10 (FERREIRA *et al*, 1999)

$$VE(\%) = ((s_{ij}^2) / \sum_{i=1}^n) \cdot 100 \quad (10)$$

Na maioria dos casos antes de se fazer propriamente a análise das componentes principais, é necessário um pré-tratamento dos dados, que nada mais é que organizar as informações na forma de uma matriz, **X**,

chamada de matriz de dados, onde as linhas representam as  $n$  amostras e as colunas representam as  $m$  variáveis. Assim, cada amostra é representada por um vetor coluna.

Com os dados organizados em forma de uma matriz,  $\mathbf{X}$ , é necessária a realização de uma manipulação prévia da matriz, para que sejam removidas as prováveis fontes de variação, pois, na maioria dos casos as variações possuem dimensões e amplitudes diferentes. Esse tratamento prévio tem por objetivo expressar cada observação em dimensão e amplitude equivalentes, sem perda de informações.

Existem diversas formas de realizar esse pré-tratamento, sendo os mais comuns: Centrar os dados na média e Auto-escalar os dados. Centrar os dados na média é um tipo de pré-processamento que consiste em calcular a média dos dados de determinada coluna, em seguida subtraí-la de cada dado da referida coluna, como na equação 11 que se segue:

$$\mathbf{X}_{cm}(ij) = \mathbf{X}_{ij} - \mathbf{X}_j \quad (11)$$

Onde:

$\mathbf{X}_{cm}(ij)$  = valor centrado na média para a variável  $j$  na amostra  $i$ ;

$\mathbf{X}_{ij}$  = valor da variável  $j$  na amostra  $i$ ;

$\mathbf{X}_j$  = média dos valores das amostras na coluna  $j$ .

Esse tipo de pré-processamento é utilizado visando filtrar possíveis ruídos. Assim, as médias das variáveis são igualadas a zero.

Já no auto-escalamento, as médias e os desvios padrões são calculados, em seguida, subtraem-se cada valor da média e divide-se o resultado pelo desvio padrão. Dessa forma, a média de cada variável será igual a zero e o desvio padrão igual a mais ou menos um. Este tipo de pré-processamento é utilizado quando se quer dar o mesmo peso para todas as variáveis, independente de sua dimensão. O auto-escalamento é representado pela equação (12) abaixo:

$$\mathbf{X}_{ij}(es) = \mathbf{X}_{ij} / S_j \quad (12)$$

Onde

$X_{ij}$ (es)= valor auto-escalonado da variável  $j$  para a amostra  $i$ ;

$X_j$ = média dos valores das amostras na coluna  $j$ ;

$S_i$ = desvio padrão dos valores da variável  $j$ .

## 2.4-ANÁLISE DE IMAGENS DIGITAIS

Quando operações levam a formação de nova imagem, a transformação é chamada de processamento de imagem. Se as operações levam a redução dos dados para chegar-se a conclusões sobre o sistema, dá-se o nome de análise de imagens. Por exemplo, em imagens médicas o resultado final da análise não é uma imagem e sim o diagnóstico do mal que aflige o paciente.

As técnicas de análise de imagens permitem a extração de informação quantitativa ou qualitativa de imagens, isto é, permitem gerar dados a partir de imagens. Nessa técnica é realizada a decomposição da imagem colorida em imagens de cores primárias, gerando através dessa, três imagens que serão processadas. Diversos métodos quimiométricos têm sido empregados na análise de imagens multivariadas, uma área da quimiometria conhecida pela sigla MIA, e que tem origem no nome em inglês para *Multivariate Image Analysis* (ESBENSEN e GELADI, 1989, GELADI *et al.* 1989; GELADI *et al.* 1992, GELADI e GRAHAN, 1996). A maioria dos métodos de MIA é baseada no uso da análise de componentes principais (PCA) das imagens após estas serem submetidas a um processo de desdobramento ao longo do modo apropriado. Este tipo de abordagem, em geral, é aplicado a uma imagem multivariada formada por diversos canais, e a análise utilizando PCA será útil na detecção de variações ao longo da imagem, e na identificação de suas causas por meio da análise dos *loadings* e *scores*.

Um processamento utilizado nas imagens multivariadas é o escalamento linear ou então o escalamento não linear. O escalamento é conhecido como operações de pré-tratamento dos dados. Exemplos de

escalamento linear são como já citado: centrar na média, escalamento pelo desvio padrão, escalamento pela variância. O escalamento não linear geralmente é aplicado quando se tem um modelo físico que apresenta resposta não linear ou quando a imagem apresenta um histograma assimétrico. Nesses casos, a aplicação de um logaritmo, raiz quadrada ou outros expoentes, podem levar a resultados visuais mais objetivos. (GELALD & GRANHAN, 1996).

Uma das ferramentas mais importantes no estudo de imagens multivariadas é a análise de componentes principais (PCA). O cálculo do PCA de uma matriz de imagens 2D, é feito calculando-se os autovalores e os autovetores das matriz que nos fornecerá os *scores* e os *loadings* da matriz. Os *loadings* são calculados a partir do produto cruzado da matriz  $X$  pela inversa desta ( $X'$ ) gerando uma nova matriz que poderemos denominar de  $Z$ , essa é decomposta no produto da matriz de *loadings* ( $P$ ) e uma matriz diagonal ( $D$ ). Os *scores* da PCA são, assim, calculados a partir dos *loadings* e a matriz reorganizada  $X$ . (LAQQA, 2007).

O número de publicações científicas relativas ao uso de imagens para obtenção de dados qualitativos e quantitativos tem evoluído nos últimos anos. A maior quantidade de trabalhos utilizando imagens em química é observado na área de microscopia, sendo as maiores aplicações em análise de superfícies. Existem também muitos trabalhos na área de biologia e bioquímica. (BORIN, 2007).

Em nosso país há, no momento, uma ampla faixa de pesquisas utilizando imagens nas mais diversas áreas dentre algumas se podem citar: avaliação de defeitos em semicondutores, medida de distribuição de ingrediente ativo em comprimidos farmacêuticos, (MOURA, 2005) detecção de traços contaminantes em células de combustível, imagens de satélites, imagens médicas (clínicas), quantificação de microorganismos em leite fermentado, determinação da cor de mel de abelhas, visão de desgaste de equipamentos utilizando análise de imagens etc. (DESCHAMPS, 2004).

### 3. METODOLOGIA

#### 3.1- AMOSTRAGEM

Foram adquiridas cervejas, em latas de 350 mL, em comércio local de acordo com o grau de aceitabilidade e da disponibilidade em mercado na cidade de Goiânia/GO no decorrer do primeiro semestre do ano de 2006. Para essa amostragem as imagens foram capturadas em formato JPEG, sendo adquiridas dez diferentes marcas de cerveja onde para cada marca foram selecionadas cinco amostras do mesmo lote de fabricação, com um universo amostral de cinquenta amostras.

No primeiro semestre de 2007 foi feita nova amostragem para analisar imagens em formato *bitmap* (BMP), sendo utilizados os mesmos critérios adotados na amostragem anterior. Foram escolhidas oito marcas de cerveja, totalizando quarenta amostras, das quais sete marcas de cervejas coincidiram com as utilizadas para a análise em formato JPEG.

#### 3.2- EQUIPAMENTOS

- a) Ultra-som Thorton T 14;
- b) Placa de petri;
- c) Béqueres de 200, 250 e 500mL;
- d) Scanner Genius Color Page-Vivid 1200 XE;
- e) Pipeta volumétrica de 25mL;
- f) Microcomputador com sistema operacional Windows<sup>®</sup> 98 ou mais atual;
- g) Software Scilab<sup>®</sup> (Copyright 1989 – 2005. SIP ENPC” and “Scilab is a trademark of SIP”; disponível em <http://siptoolbox.sourceforge.net/>).

### 3.3- PROCEDIMENTO EXPERIMENTAL

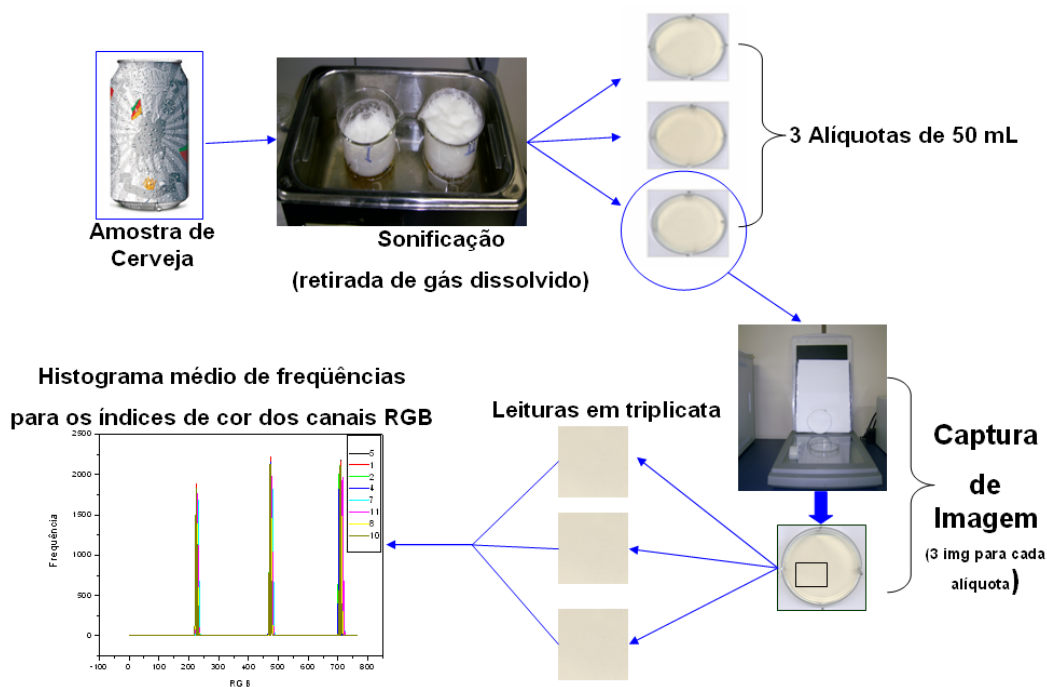
#### 3.3.1- CAPTURA DAS IMAGENS DE CERVEJAS

Inicialmente foram analisadas 10 marcas de cervejas em lata de 350mL, através de imagens capturadas em formato JPEG com resolução de 300 ppp. Para a análise, retirou-se de cada lata de cerveja, na temperatura ambiente, cerca de 200 mL da amostra que foi transferida para um béquer e colocada em ultra-som por cerca de 30 minutos, até não haver desprendimento de gás. Da cerveja desgaseificada, retirou-se uma alíquota de 50 mL e transferiu para a placa de petri, de maneira que não houvesse formação de bolhas (dissolução de gás) na cerveja. A placa de petri com cerveja desgaseificada foi colocada sobre a mesa do *scanner*, em uma posição marcada no equipamento. Para a captura de imagem no scanner foi fixado no aplicativo Assistente e scanner e câmera, do Sistema Operacional *Windows* na opção personalizar as configurações de contraste e brilho em 64 e 25 respectivamente. Foram retiradas três alíquotas de cada amostra de cerveja, conforme descrito anteriormente, e de cada alíquota foram capturadas quatro imagens – onde a primeira imagem era sempre desprezada por tratar-se da fase de aquecimento do *scanner* – sendo armazenadas três imagens de cada alíquota. Das imagens armazenadas definiu-se uma região homogênea, idêntica, quadrada e com dimensão de 200x200, conforme Figura 4.

Foi realizado o procedimento descrito anteriormente para imagens digitalizadas de cervejas com resolução de 600 ppp, formato JPEG.

Foi realizado experimento para observar a alteração de cor da cerveja quando exposta a luz e ao ar para tanto, escolheu-se uma amostra a qual após o processo de desgaseificação foi retirada uma alíquota da qual se capturaram três imagens. Essa alíquota permaneceu na placa de petri sobre a mesa de scanner por 24 horas, sem proteção física, sendo capturadas imagens hora a hora durante as quatro horas e após vinte e quatro horas houve nova captura de imagens.

Para a nova amostragem o procedimento foi idêntico ao descrito porém fixou-se a resolução em 300 ppp e formato BMP.

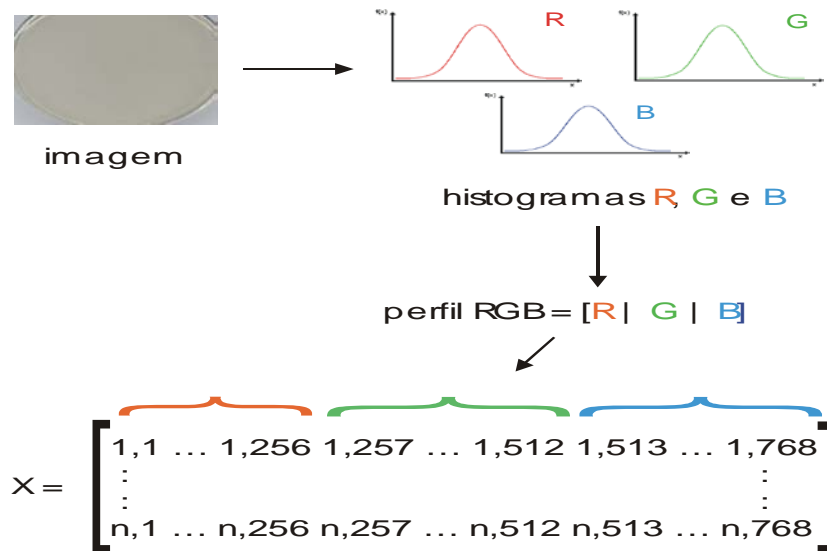


**Figura 4-** Fluxograma para captura de imagens de cervejas.

### 3.3.2- ANÁLISE DAS IMAGENS DIGITALIZADAS

Para a etapa da análise das imagens digitais foi desenvolvido um algoritmo conforme o Anexo A para extrair histogramas de freqüências dos índices de cor dos canais do sistema RGB utilizando o aplicativo matricial *Scilab* com o pacote *SIP*. Para cada imagem digitalizada foi obtido uma matriz transposta  $Z_{1 \times 768}$  sendo as informações da coluna 1 a 256 referentes ao canal R, coluna 257 a 512 canal G e coluna 513 a 768 ao canal B. A partir da matriz  $Z$  gerada para cada imagem foi feita uma matriz  $W$  contendo a média dos histogramas para cada marca. Criou-se uma matriz  $X$ , definida como matriz de dados originais, agrupando todas as matrizes médias de

freqüências de cada marca, com as marcas dispostas nas linhas e os índices R, G e B em colunas, conforme Figura 5.



**Figura 5** - Fluxograma para obtenção da matriz X

### 3.3.3-Análise de Componentes Principais (PCA) das imagens digitais

A ferramenta multivariada PCA foi empregada para a classificação das imagens. PCA consiste em uma mudança de base dos dados multivariados, representados pela matriz **X**, a essa matriz foi aplicado um algoritmo (FERREIRA, 1999) que pode ser dividido em três etapas, conforme Anexo B.

A primeira etapa do algoritmo consiste no pré-processamento dos dados, onde os dados da matriz original (matriz X) são centrados na média. A segunda etapa a matriz com os dados centrados na média ( $X_{cm}$ ) foi decomposta em valores singulares obtendo a matriz de *loadings* e *scores*. A terceira etapa consiste na determinação da variância explicada por cada autovetor ou componente principal.



## 4- Resultados e Discussão

### 4.1 – Análises das imagens

Na primeira parte deste trabalho foram analisadas imagens de dez diferentes marcas de cerveja, na qual as imagens foram digitalizadas no formato JPEG. As cervejas utilizadas foram adquiridas no comércio local e analisadas dentro dos prazos de validade de cada marca. De cada marca cerveja era retirada uma amostra, de cerca de 200 mL, e colocados em um béquer posteriormente levados a um banho de ultra-som, com o propósito de se retirar o gás dissolvido na bebida, de modo a evitar interferências nas imagens analisadas, como o aparecimento de bolhas, descaracterizando o padrão de cor da imagem da amostra. Não foi estipulado um tempo exato para que cada amostra permanecesse em ultra-som, mas foi percebido que o tempo médio foi de vinte a trinta minutos, tempo que poderia variar para cada marca, porém nenhuma das amostras excedeu este tempo. Para cada marca foram retiradas três alíquotas de cada amostra da qual foram capturadas quatro imagens, sendo a primeira leitura desprezada, por se tratar de uma imagem visivelmente diferente das demais, considerada um *outlier*, ou imagem anômala, obtida durante a fase de aquecimento do *scanner*.

Durante os ensaios iniciais de captura das imagens de cervejas pôde ser observado que as primeiras imagens vistas no monitor do computador não apresentavam uma cor que lembrasse diretamente a cor natural da cerveja, pois apresentavam tom amarelo acinzentado. Essas imagens passaram por alguns ajustes de brilho e contraste, por tentativa e erro, onde se buscou um tom visualmente perceptivo que mais se assemelhasse à cor natural da cerveja. Para tanto foram fixados os seguintes atributos de cor: 25 para o brilho e 64 para o contraste.

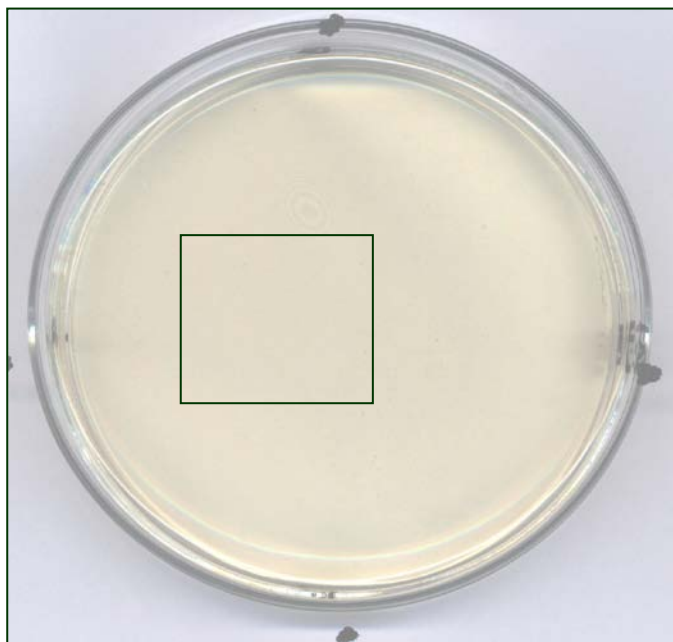
Como forma de melhorar a reflexão da fonte de luz do *scanner* e gerar imagens ainda mais próximas das vistas a “olho nu”, foram realizadas algumas adaptações na tampa da placa de petri e no anteparo do scanner,

conforme a Figura 6, onde é possível se observar os anteparos brancos na tampa da placa de petri e no anteparo da mesa do *scanner*. Com essas adaptações obteve-se uma maior reflexão da fonte de luz, o que resultou em imagens com melhor aparência e nitidez.



**Figura 6-** *scanner* e placa de petri adaptados com anteparos brancos.

Após a captura das imagens de todas as marcas de cerveja, foi delimitada uma região retangular interna na placa de petri definida conforme Figura 7, da qual se analisaram todas as imagens. Essa região retangular é idêntica para todas as análises e é dessa região da imagem que se obtém a matriz de dados da qual se tem as informações da imagem digital capturada.



**Figura 7** - Imagem de cerveja contendo região retangular definida para análise.

Na mesa de captura de imagem do *scanner* foi definido o local onde a placa de petri era colocada para proceder à digitalização da imagem. Em todas as análises foi utilizada uma única placa de petri, a qual foi devidamente marcada e colocada na mesma posição no *scanner* em todas as capturas das imagens.

Definidos os padrões das etapas da análise, quarenta e cinco imagens para cada marca foram digitalizadas. As imagens obtidas foram processadas pelo software Scilab, o qual foi escolhido por oferecer um ambiente amigável, semelhante ao MatLab, porém de domínio público. O perfil de cor das imagens digitais foi avaliado para cada um dos canais do sistema RGB, sendo obtida uma matriz coluna,  $Z_{768 \times 1}$ , que representa o número de vezes que um determinado índice de cor está contido na área digitalizada. Como para cada marca foram obtidas quarenta e cinco imagens, criou-se uma matriz  $Y_{768 \times 45}$  da qual se fez uma média e obteve-se uma matriz média  $W_{768 \times 1}$  para cada marca. Todo esse processamento foi realizado para as dez marcas e a partir das matrizes coluna médias transpostas gerou-se uma nova matriz  $X_{10 \times 768}$ , que foi definida como sendo a matriz de dados originais. A matriz  $X$  contém todas as informações que se necessitam para análise de cor relacionadas com as marcas de cerveja,

dispostas em linhas e com as variáveis, índices de cores dos canais R,G e B, dispostos em colunas de modo que as informações são distribuídas da seguinte forma: a) coluna 1 a 256, canal R; b) coluna 257 a 512, canal G; e c) coluna 513 a 768, canal B.

A partir da matriz **X** os dados são pré-processados (centrados na média) e por meio do Scilab a matriz bidimensional foi decomposta nos vetores de *loadings* e *scores*. Para a decomposição da matriz de dados originais, fez-se uso do método de decomposição em valores singulares (**SVD**), um algoritmo matemático que possibilita determinar auto-vetores e autovalores. Para a produção dos gráficos de PCA e dos histogramas médios das frequências também se utilizou do programa matemático Scilab. Cabe ressaltar que qualquer pacote computacional para elaboração de gráficos pode ser utilizado para esse fim, sendo que no presente trabalho optou-se pelo Scilab, pelo fato de ser um software gratuito e poder ser utilizado como pacote dedicado a quimiometria, tendo seu uso viável em todas as etapas do trabalho.

Para a captura das imagens foi utilizado o programa Assitente de Scanner e Câmera, do pacote Microsoft Windows®, o qual possibilitou que as imagens fossem capturadas com as seguintes resoluções: 75 ppp, 100 ppp, 150 ppp, 200 ppp, 300 ppp e 600 ppp. Dentre essas foram escolhidas ao acaso, inicialmente, as resoluções de 100 ppp, 300 ppp e 600 ppp. Nesse ponto definiu-se que quando se tratar das resoluções das imagens será utilizada a terminologia ppp (ponto por polegada), que é uma tradução direta do termo dpi, que significa *dots per inch*.

As primeiras análises das imagens, com resolução em 100 ppp se mostraram rápidas e duraram pouco mais de cinco segundos para cada imagem capturada, o que se atribuiu à baixa resolução utilizada, que gerou arquivos pequenos, com imagens de baixa qualidade. Nas análises das imagens em 300 ppp o tempo gasto para captura de cada imagem foi pouco maior que o gasto para imagens com 100 ppp, algo em torno de 5 a 8 segundos, gerando imagens com boa nitidez e aparência, diferentemente das imagens com a menor resolução. Nas análises em 600 ppp o tempo

gasto foi cerca de duas vezes maior que o gasto em imagens de 300 ppp, gerando imagens “pesadas” o que requer grande disponibilidade de espaço livre em disco para armazenamento das imagens. Isso se torna um dos entraves para escolha dessa resolução como um padrão para as análises.

Durante as análises das imagens sensíveis alterações de cor foram percebidas nas amostras de cerveja quando expostas à luz e ao ar. Apesar desse fenômeno não ser o objetivo principal do presente trabalho, esse fenômeno passou a ser avaliado como sendo um objetivo secundário, Escolheu-se uma amostra da marca 1 e após o processo de degaseificação retirou-se uma alíquota da qual capturou-se três imagens. Essa alíquota permaneceu na placa de petri sobre a mesa de scanner por 24 horas, sem proteção física, sendo capturadas imagens hora a hora durante as quatro horas e após vinte e quatro horas houve nova captura de imagens. A Figura 8 apresenta as imagens representativas do experimento de alteração de cor.



**Figura 8** - Imagem representativas da marca 1 em diferentes intervalos de tempo (0 h, 1 h, 2 h, 3 h, 4 h e 24 h) quando exposta ao ar.

O motivo que justifica as alterações de cor deve-se, provavelmente, ao processo de escurecimento não-enzimático, ou mesmo à oxidação natural dos componentes da bebida quando expostos à luz e ao ar.

Na segunda parte deste trabalho, com o intuito de ampliar o estudo da análise das imagens, fez-se uso de outro formato para arquivos de imagem, o formato *bitmap*, BMP. Esse formato, amplamente difundido para imagens, gera arquivos maiores que o formato JPEG, o que se deve ao fato de não comprimir os dados das imagens.

Para as análises das imagens no formato BMP, foi realizada uma nova amostragem das marcas das cervejas as quais foram escolhidas de acordo com suas disponibilidades no mercado na região metropolitana de

Goiânia. Nessa nova amostragem foram escolhidas oito marcas, dentre as quais, sete coincidiram com as marcas utilizadas na análise das imagens com formato JPEG e apenas uma marca diferente. A Tabela 3 apresenta algumas características das marcas com suas denominações atribuídas neste trabalho, uma vez que por motivos legais seus nomes comerciais foram omitidos. No Anexo C a nova marca de cerveja adicionada é a denominada marca 11 e as outras marcas utilizadas na análise com imagens no formato JPEG seguem com denominação de 1 a 10. As sete marcas coincidentes em ambas análises das imagens são as seguintes: 1, 2, 4, 5, 7, 8 e 10.

O resultado das análises das imagens nos formatos BMP e JPEG permitiram classificar as marcas de cervejas, mas, por se tratar de formatos e arquivos de tamanhos diferentes e por ainda conter um diferente conjunto de dados os resultados apresentados em ambos não devem ser diretamente comparados mas para classificação de cervejas ambos podem ser utilizados independentemente.

Os resultados da análise das imagens são apresentados através de três tipos de gráficos: histogramas médios das frequências dos tons do sistema RGB e PCA (gráfico dos *scores* e gráfico dos *loadings*). Os histogramas médios das frequências, possibilita uma avaliação simultânea das frequências nos três canais do sistema *true color*, onde os canais são inseridos de forma subsequente, de modo que sejam analisadas as 768 variáveis, as quais foram divididas da seguinte forma: canal R, variáveis de 1 a 256, canal G de 257 a 512 e B de 513 a 768. Com essa disposição do histograma médio das frequências é possível avaliar o perfil RGB de cada marca. Os histogramas médios das frequências possibilitam estimar os intervalos de índices mais importantes para a classificação das marcas de cerveja.

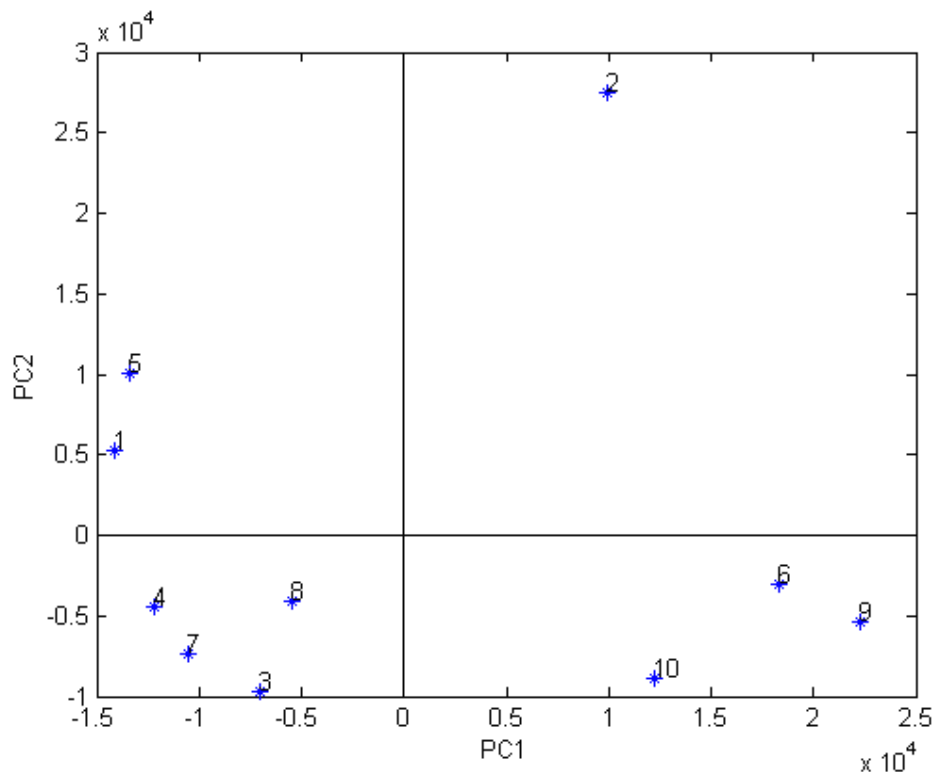
Com o gráfico dos *loadings* é possível confirmar os resultados obtidos com o histograma médio das frequências, uma vez que esse gráfico apresenta as variáveis de maiores pesos, dentre as 768 do sistema RGB. Ainda a partir do gráfico dos *loadings* é possível observar o comportamento das variáveis, as quais de acordo com o posicionamento das marcas no

gráfico dos *scores* indicam as marcas que tendem a ter tons mais claros ou mais escuros, avaliando o comportamento dos índices de cor no gráfico dos pesos. A análise dos pesos para as PCs fica muito prejudicada com grande quantidade de índices com valor zero ou próximo de zero (variáveis de menor importância), isto dificulta a visualização e distribuição dos índices importantes. Em função destes aspectos, os índices de menor importância foram sistematicamente retirados para facilitar a interpretação do gráfico dos *loadings*.

O gráfico dos *scores* possibilita avaliar o comportamento entre as marcas de cervejas, evidenciando as marcas similares e diferentes de acordo com a região em que estão inseridas na PC. O gráfico dos *scores* é a projeção das amostras na direção de cada componente principal. Na maior parte dos casos é possível observar a formação de agrupamentos que permitem classificar as marcas por grupos de similaridades.

Os resultados apresentados pelo histograma médio de frequência e pela PCA (gráficos dos *loadings* e dos *scores*) são complementares. Por esse motivo, recomenda-se a utilização de ambos, uma vez que isso aumenta a consistência das interpretações.

## 4.2 – Análise das imagens com resolução 300 ppp no formato JPEG

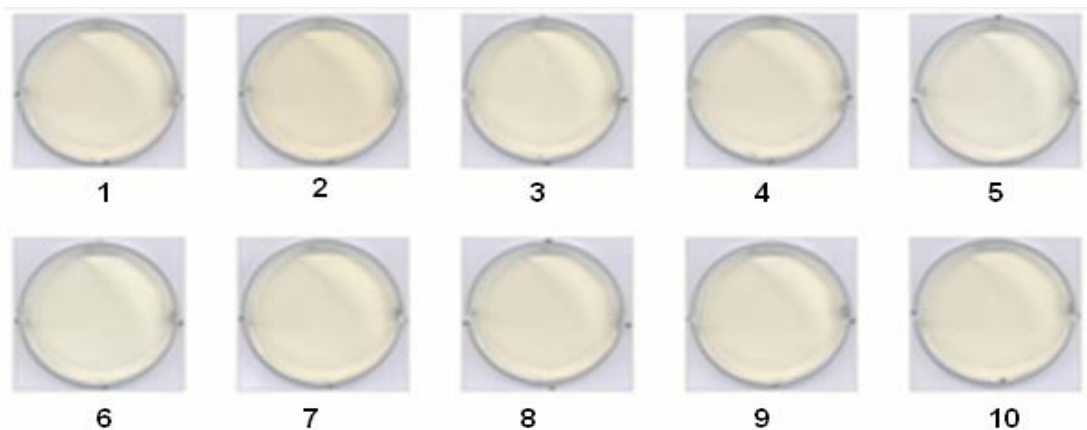


**Figura 9** – Gráfico dos *scores* das duas primeiras componentes principais, PC1 e PC2, que explicam, respectivamente, 44,16% e 29,17% das variâncias dos canais RGB com 73,33% da variância acumulada das marcas 1 a 10 – resolução 300 ppp, no formato JPEG.

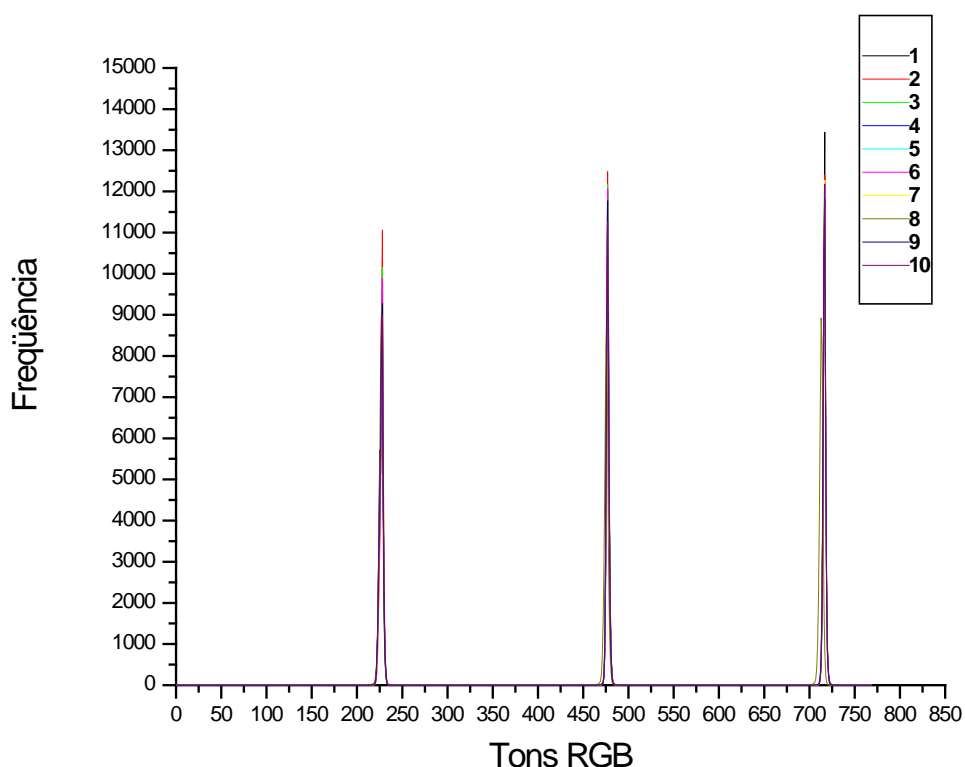
A partir da Figura 9 é possível observar a formação de quatro grupos distintos de marcas de cervejas, com uma clara separação nos quadrantes. O primeiro quadrante contém as marcas 1 e 5. O segundo quadrante contém apenas a marca 2, separada das demais marcas. O terceiro quadrante contém a maior quantidade de marcas, são elas as marcas: 3, 4, 7 e 8, sendo as marcas que possuem maior similaridade uma vez que são as que apresentam melhor agrupamento. O quarto quadrante contém as demais marcas, são elas: 6, 9 e 10. Com a avaliação do gráfico dos *scores* apresentado pela Figura 9, com uma variância acumulada de mais de 70% para as duas primeiras componentes principais, visualiza-se uma boa



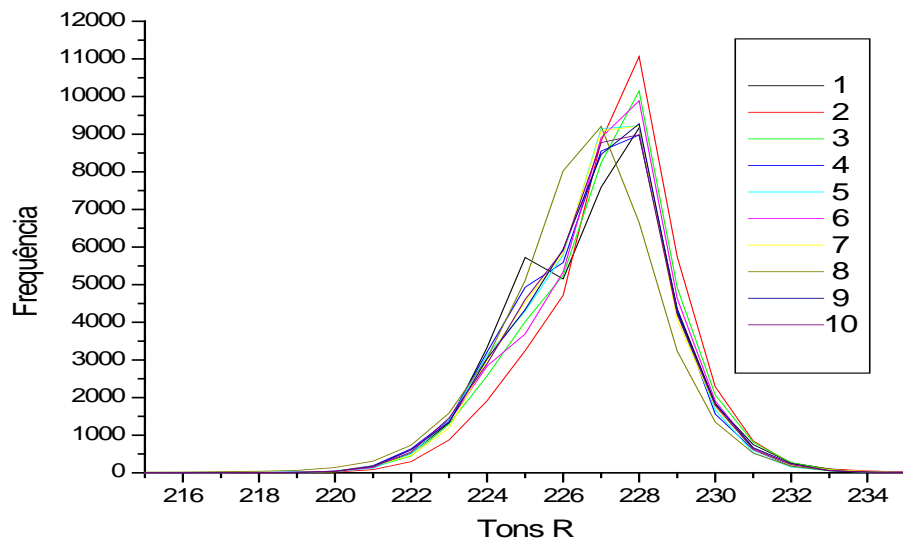
separação das marcas, onde é possível afirmar que, dentre as dez marcas avaliadas, a marca 2 é a que mais difere tendo como atributo a cor.



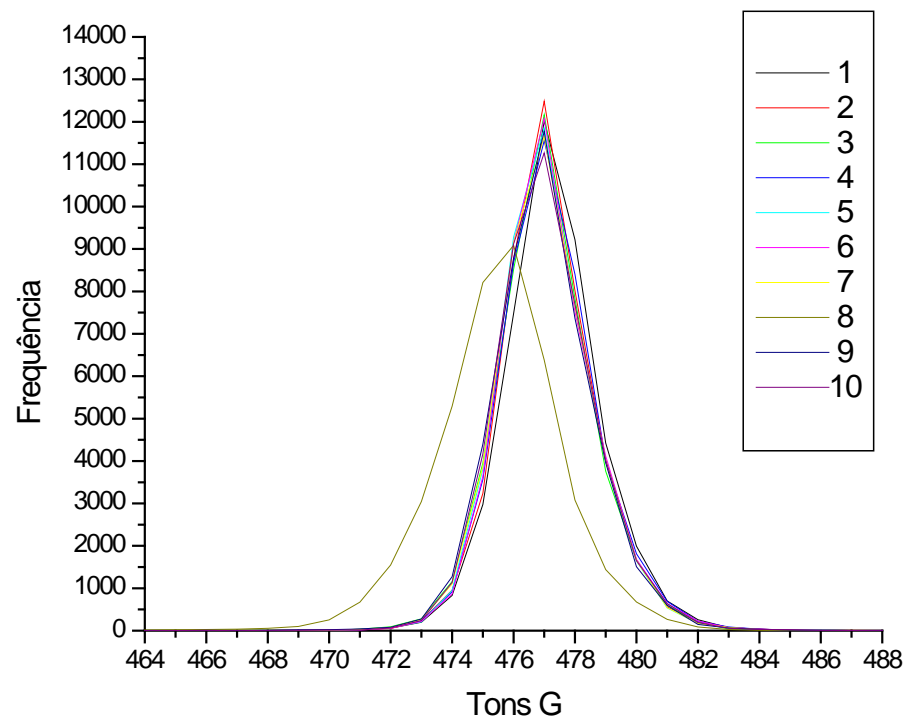
**Figura 10** - Imagens representativas das 10 diferentes amostras de cervejas com resolução de 300 ppp. Imagens das marcas 1 a 10 no formato JPEG.



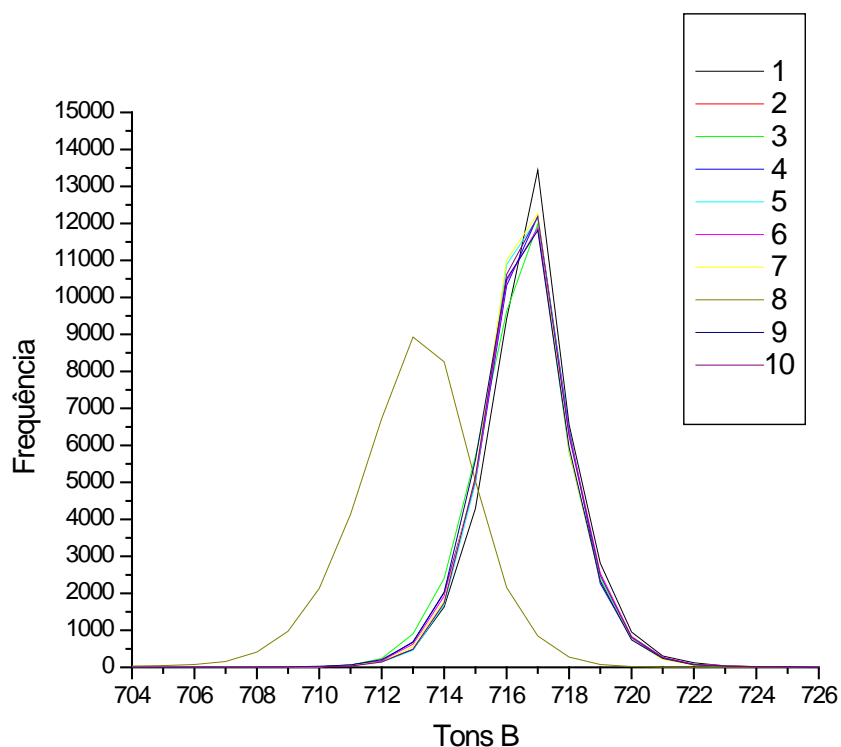
**Figura 11** – Histograma médio das frequências versus tons RGB das amostras de cervejas das marcas 1 a 10, com resolução de 300 ppp no formato JPEG.



**Figura 11.1** – Histograma médio ampliado das frequências versus tons R das amostras de cervejas das marcas 1 a 10, com resolução de 300 ppp no formato JPEG.



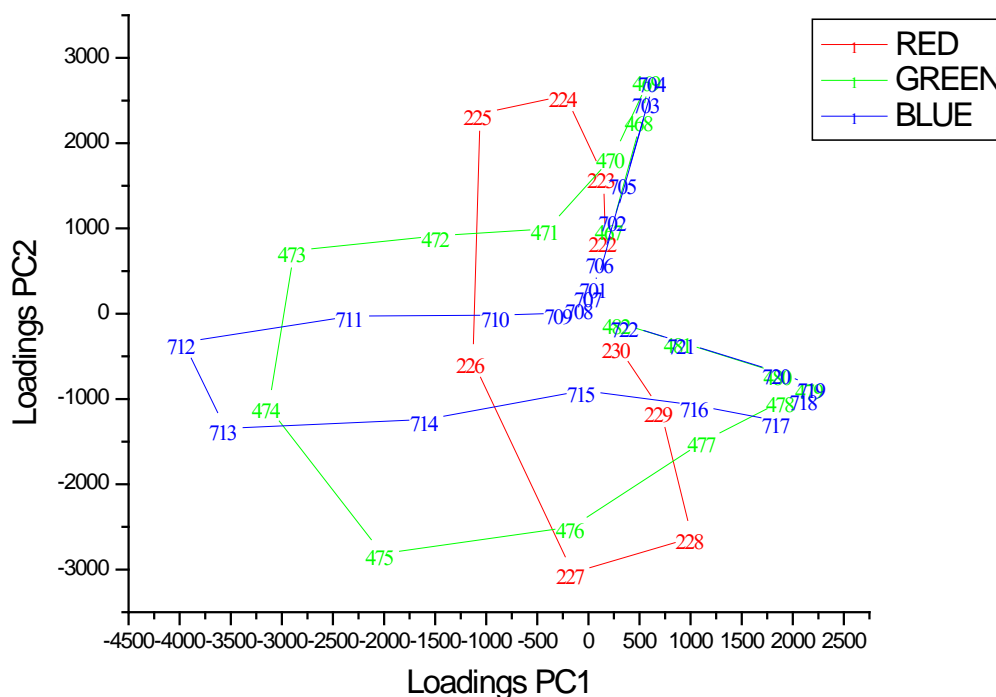
**Figura 11.2** – Histograma médio ampliado das frequências versus tons G das amostras de cervejas das marcas 1 a 10, com resolução de 300 ppp no formato JPEG.



**Figura 11.3** – Histograma médio ampliado das freqüências versus tons B das amostras de cervejas das marcas 1 a 10, com resolução de 300 ppp no formato JPEG.

Na Figura 11 é possível verificar o intervalo dos índices de cor com maiores freqüências, os quais permitem avaliar o comportamento de cada marca. O histograma médio de frequências foi desmembrado e ampliado nas figuras 11.1, 11.2 e 11.3, respectivamente para o canal R, G e B, visando fornecer uma melhor visualização dos tons de maior freqüência para cada marca. Os intervalos dos índices de maiores freqüências podem ser considerados como sendo as variáveis de maior importância, sendo eles para o canal R entre os índices 218 e 234, conforme histograma médio ampliado para o canal R, Figura 11.1, para o canal G, entre 466 e 484, conforme é explicitado na Figura 11.2, que equivalem de 194 a 244, e para o canal B entre 706 e 724, vide Figura 11.3, que equivalem 188 a 238. O histograma médio das freqüências, nos informa o comportamento dos histograma de cada marca, o que demonstra o seu perfil ou identidade. No histograma médio das freqüências representado na Figura 11, as diferenças

nas linhas de cada marca são pequenas, podendo apresentar pequenas diferenças perceptíveis nos picos das curvas, o que evidencia uma pequena distinção nas cores das cervejas amostradas, uma vez que as linhas não são totalmente coincidentes. Caso fossem, poderíamos dizer que as marcas possuíam tons similares de cor. A pequena discordância na curva indica que as marcas possuem freqüências diferentes em alguns índices de cor do sistema RGB. Complementado os resultados apresentados pelo histograma médio das freqüências, a PCA dos *loadings*, Figura 12, evidencia os índices de maior importância e que contribuem para a classificação das marcas.



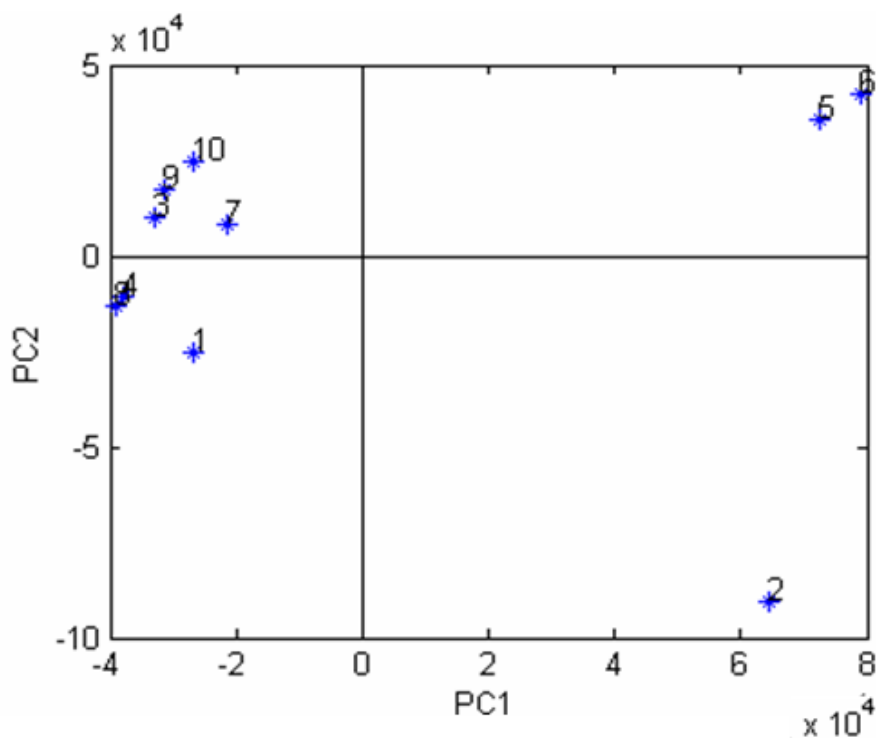
**Figura 12** - Gráfico dos *loadings* de PC1 e PC2 das marcas 1 a 10 de cervejas. Imagens digitalizadas com resolução de 300 ppp, no formato JPEG. As numerações referem-se aos valores dos índices de cor para os canais R, G e B justapostos.

Da Figura 12 é possível depreender que os índices de cor estão aumentando no sentido anti-horário, sendo possível observar as variáveis de maior peso que possibilitam a classificação das marcas: para o canal R de 222 a 230, canal G de 467 a 480 e canal B 701 a 722. Valores que

equivalem respectivamente aos índices R, G e B: 221 a 229, 212 a 226 e 190 a 210. Caso se prefira escolher apenas um canal para a classificação das marcas, infere-se pelo gráfico dos *loadings* que o canal B é o que apresenta uma melhor amplitude para separação, ou seja, apresenta mais variáveis que contribuem para uma classificação das marcas, o que pode ser certificado avaliando as discrepâncias das curvas no histograma médio das freqüências para o canal B na Figura 11.

Comparando-se a posição das amostras no gráfico dos *scores*, Figura 9, e o comportamento dos índices de cor na Figura 12, é possível inferir que a marca 2 é a mais escura dentre as avaliadas uma vez sua posição pelo gráfico dos *scores* coincide com a região de menores índices e por sua vez a marca 6 é a mais clara por estar em região dos índices maiores. Para a avaliação deve-se notar o aumento dos índices no sentido anti-horário tendo como referência inicial o índice da marca 2 que aumenta até a marca 6.

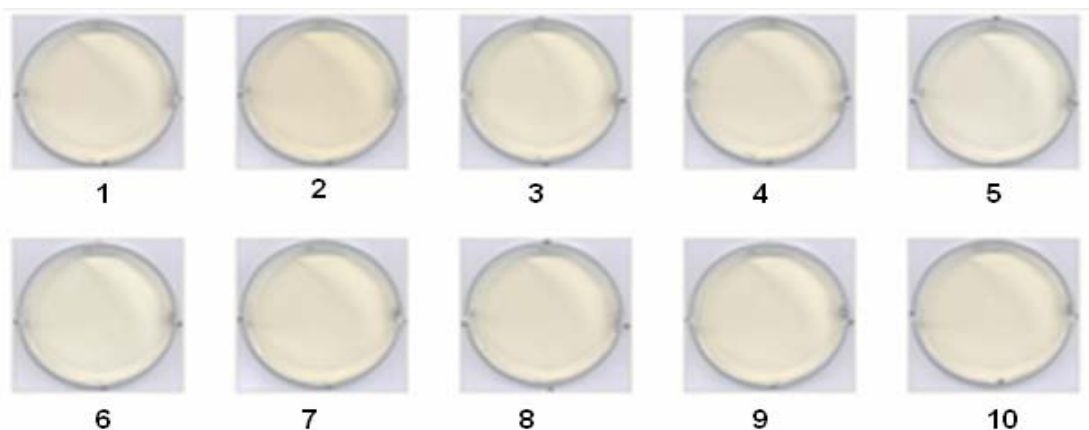
#### 4.3 - Análise das imagens com resolução 600 ppp no formato JPEG



**Figura 13** – Gráfico dos *scores* PC1 e PC2, que explicam, respectivamente, 50,53% e 29,82% das variâncias dos canais RGB com 80,35% da variância acumulada das marcas 1 a 10, com resolução 600 ppp no formato JPEG.

A Figura 13, apresenta o gráfico dos scores para as dez marcas de cervejas avaliadas no formato JPEG, com resolução de 600 ppp, do qual é possível observar a formação de três grupos distintos, sendo um composto pelas marcas 5 e 6, outro onde há apenas a marca 2, cuja análise das imagens com resolução de 300 ppp já se mostrou como uma marca com característica de tons cores diferentes das demais marcas, e um grande agrupamento composto pelas outras marcas. Esse gráfico dos scores apresenta uma alta variância acumulada para as duas primeiras componentes principais, variância maior que 80%.

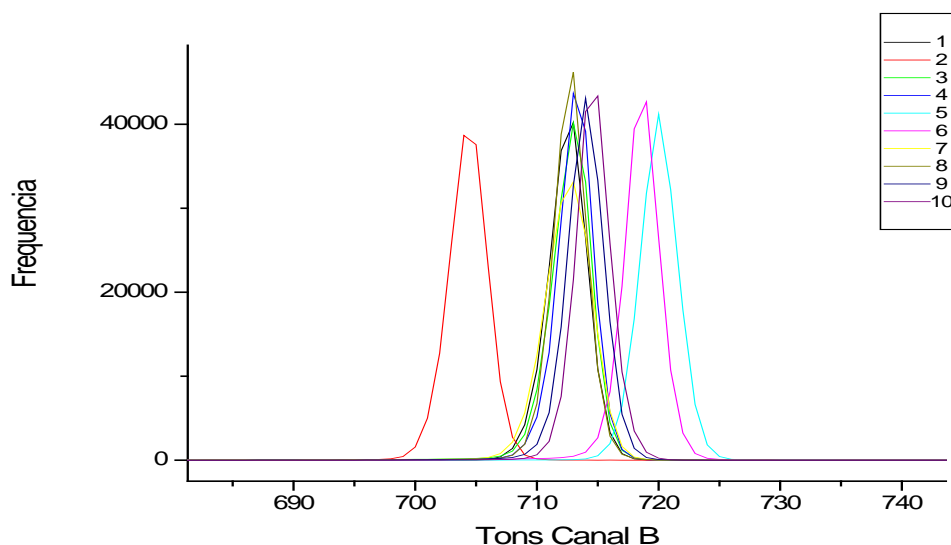
Apesar das imagens com resolução de 600 ppp no formato JPEG serem mais nítidas que as de 300 ppp de resolução, é percebido uma separação similar àquela obtida para a resolução de 300 ppp. A variância acumulada pelas duas primeiras componentes principais para a análise com resolução de 600 ppp foi de cerca de 7 % maior que a PCA com 300 ppp, porém o tempo gasto e a quantidade de espaço de memória necessários foi maior, o que não justifica preterir a utilização da resolução de 300 ppp. Ressalta-se que o objetivo é de classificar marcas de cervejas a partir de análise das imagens tendo como atributo a cor, fazendo uso das imagens capturadas por *scanner* de mesa, de modo que a análise seja rápida e acessível. Assim, considerando os recursos disponíveis neste trabalho, será padronizada o uso das imagens com resolução de 300 ppp, pois essa resolução permite uma boa classificação das marcas de cerveja com celeridade.



**Figura 14** - Imagens representativas das 10 diferentes amostras de cervejas com resolução 600 ppp no formato JPEG.

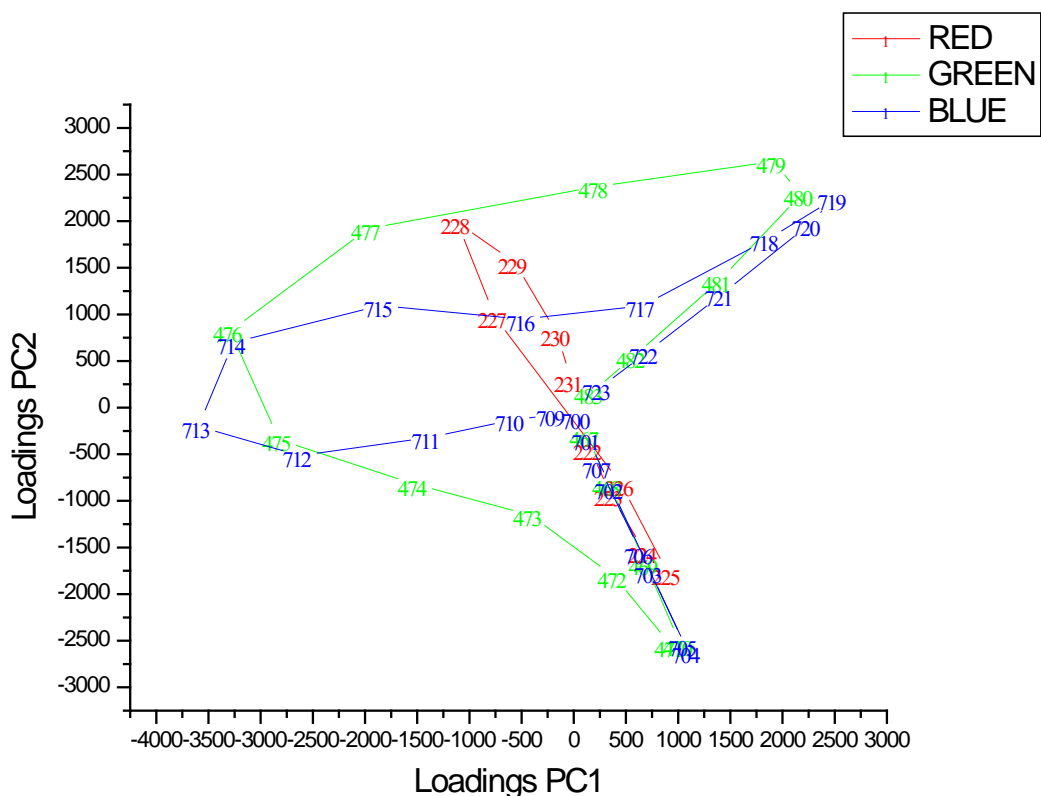
A Figura 14 é uma imagem representativa das dez marcas de cervejas analisadas com resolução 600 ppp no formato JPEG.

Uma vez que nas análises das imagens com resolução de 300 ppp no formato JPEG foi percebido que um dos canais apresentava uma melhor amplitude de separação, qual seja o canal B. Para resolução de 600 ppp e no formato JPEG os histogramas médios das freqüências para as dez marcas de cervejas foram gerados e os histogramas para o canal B constam da Figura 15. Nessa figura é clara a separação de três marcas, a marca 2 para índices menores e respectivamente as marcas 5 e 6 para índices maiores. Com esses histogramas é possível chegar a resultados similares aos obtidos para a análise com resolução de 300 ppp onde a marca 2 foi considerada como sendo a marca mais escura e a marca 6 a mais clara – por terem maiores freqüências para índices menores e maiores, respectivamente, após a comparação entre os gráficos dos *loadings* e dos *scores*. Da figura 15 observa-se que os índices de cor para o canal B que permitem a classificação das marcas é o intervalo de 700 a 725, que referem-se aos índices de 187 a 212. Para a marca 2 é possível perceber que o intervalo das freqüências dos índices é de 700 a 710, que equivalem aos índices de 187 a 197.



**Figura 15** – Histograma médio das freqüências versus tons Canal B de amostras das marcas 1 a 10 de cervejas, com resolução de 600 ppp no formato JPEG

Depreende-se dos histogramas que as marcas de referências para maiores índices, 5 e 6, e menores índices, 2, não coincidem os seus intervalos de índices de cor para a classificação, de modo que é claro que essas marcas possuem perfis diferentes.



**Figura 16** – Gráfico dos *loadings* de PC1 e PC2 das marcas 1 a 10 de cervejas resolução de 600ppp, no formato JPEG. As numerações referem-se aos valores dos índices de cor para os três canais justapostos.

A Figura 16 apresenta a PCA de *loadings* para as marcas de cervejas com resolução de 600 ppp no formato JPEG onde é possível observar que os índices de cor aumentam no sentido anti-horário. Avaliando apenas no canal B percebe-se que há um aumento no número do índice de cor que se inicia em 700 e vai até o índice 725, de modo que pela comparação com a



posição das amostras nos gráficos de *scores* na Figura 13 fica mais uma vez comprovado que a amostra 2 está na região de menores índices, e as marcas 5 e 6 na posição dos índices maiores, ambos para maiores valores positivos de *loadings*.

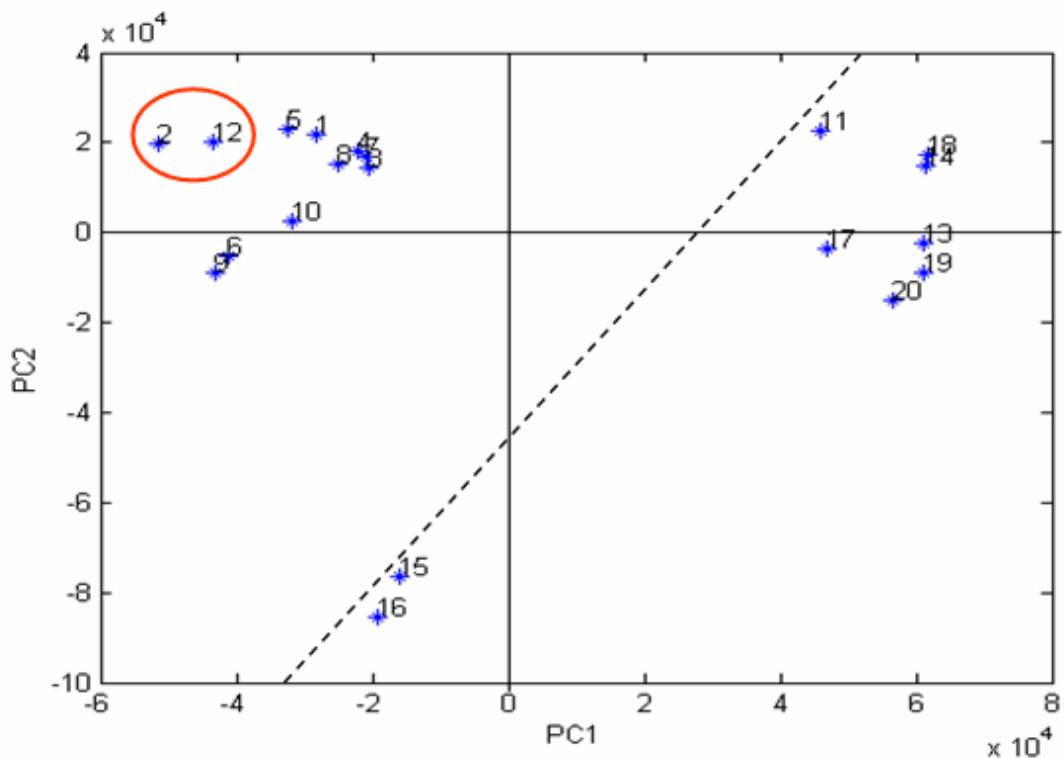
Por meio de análise visual é possível perceber tanto pelo gráfico de *scores* quanto pelo de *loadings* que o canal com maior amplitude trata-se do canal B, que favorece a análise. Por sua vez o canal R, é o de menor amplitude, sendo possível avaliar apenas os índices 222 a 231, o que dificulta inferir sobre as amostras.

#### **4.4 – Análise das imagens com resoluções 300 ppp e 600 ppp no formato JPEG**

Com o intuito de se analisar o comportamento das imagens das marcas de cervejas nas resoluções 300 ppp e 600 ppp simultaneamente, gerou-se uma PCA, conforme pode ser observado nas Figuras 17 e 18, gráfico dos *scores* e gráfico dos *loadings*, contendo as dez diferentes marcas nas duas resoluções analisadas no formato JPEG. Para melhor estabelecimento das denominações atribuídas às marcas, uma vez que o gráfico dos *scores* gerado apresentam 20 diferentes amostras, foi criada a Tabela 2.

A Figura 17 apresenta o gráfico dos *scores* das duas primeiras componentes principais com uma variância acumulada de cerca de 72%, que possibilita afirmar que há coerência do resultado. As amostras constantes na PCA apresentam posições bem características. Pode ser observado que para efeito de distinção entre as resoluções, no gráfico dos *scores* foi colocado uma linha diagonal de referência onde abaixo desta é possível perceber que estão amostras com resolução em 600 ppp, com exceção da marca 12, que corresponde à marca 2, marca essa já caracterizada como marca com tons de cor diferente quando comparada com as demais marcas, e que era de se esperar um comportamento diverso. As marcas 15 e 16, que se referem às marcas 5 e 6, encontram-se distante

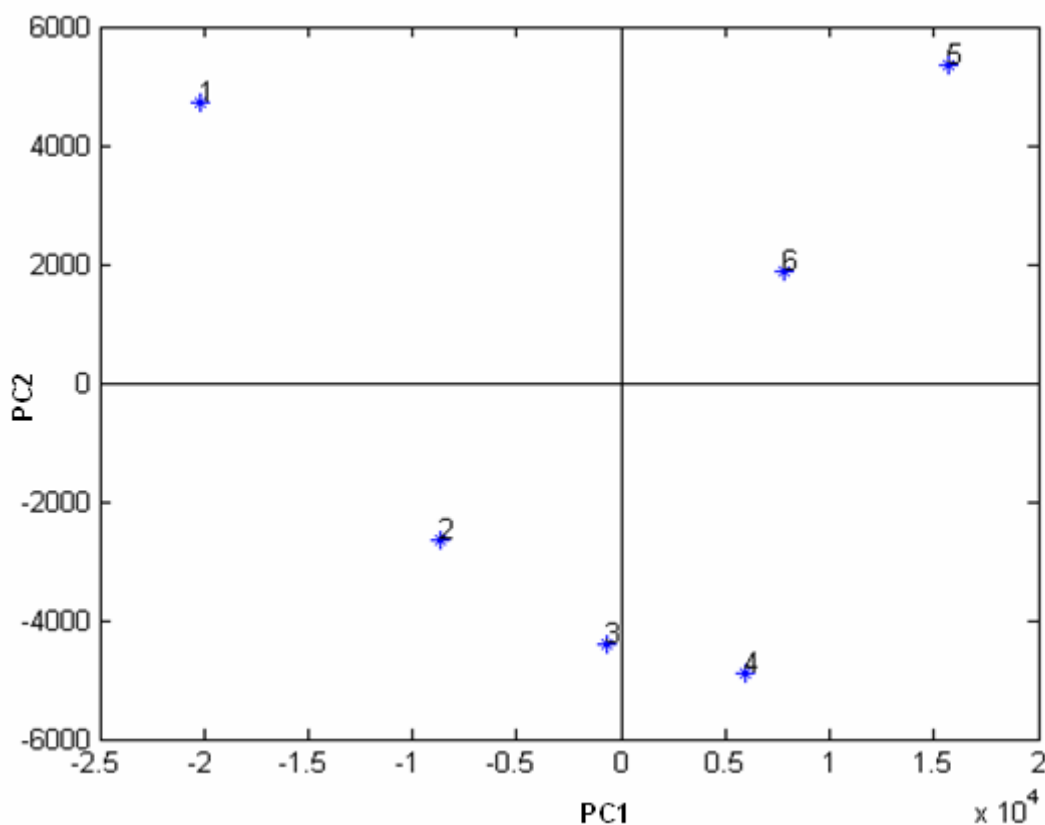
das demais, o que foi avaliado quando da análise em separado, na resolução de 600 ppp, aonde foi verificado que essas marcas eram atribuídas como referências para índices maiores, apresentando tons de cores mais claros, em oposição à marca 2 (12). As marcas que se encontram acima da linha tracejada da Figura 17 representam as amostras com resolução de 300 ppp. É perceptível pelo gráfico que com a inserção das imagens com resolução de 600 ppp, as marcas com resolução de 300 ppp apresentam um comportamento mais agrupado, mas ainda assim é possível verificar que a marca 2 encontra-se um tanto separada do aglomerado que forma as demais marcas.



**Figura 17** – Gráfico dos scores PC1 e PC2 que explicam respectivamente 48,46 % e 23,49 % da variância dos canais RGB com 71,95% da variância acumulada dos dados das 10 diferentes marcas em escalas de 300 ppp e 600 ppp no formato JPEG. A linha tracejada é indicativo da separação das imagens com diferentes resoluções à exceção das marcas 2 (12).



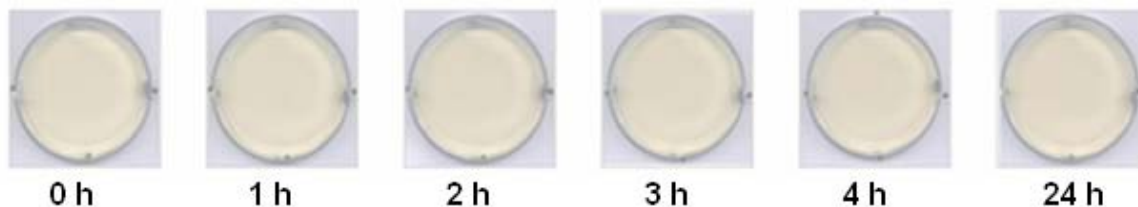
nas amostras de cervejas, escolheu-se uma das marcas a qual foi submetida à análise das imagens em diferentes intervalos de tempo, expondo a amostra à luz e ao ar. A marca escolhida foi a marca 1, a qual foi escolhida ao acaso. Dessa retirou-se uma amostra como nas demais análises anteriores e se executou todo o procedimento até a retirada de uma alíquota. Da alíquota foram capturadas triplicatas a cada hora, durante as primeiras quatro horas, e uma nova triplicata foi obtida após vinte e quatro horas de exposição. Após as capturas das imagens se executou o procedimento de análise das imagens onde, com os dados gerados pelas imagens em diferentes intervalos de tempo, foi feita a PCA. A Figura 19, representa o gráfico dos scores da análise de alteração de cor da marca 1 no formato JPEG com resolução de 300 ppp.



**Figura 19** – Gráfico dos scores PC1 e PC2, que explicam 98,98% da variância acumulada dos dados da marca 1 para diferentes intervalos 1- 0 h, 2- 1 h, 3- 2 h, 4- 3 h, 5- 4 h e 6- 24 h com resolução de 300 ppp.

Depreende-se do resultado apresentado pela Figura 19, que na primeira hora de exposição da amostra de cerveja a luz e ao ar houve significativa alteração da cor, conforme pode ser visto pela distanciação do ponto 1 dos demais. Pela figura percebe-se que da primeira à terceira hora a mudança de cor foi gradativa, pela pequena distanciação dos pontos 2, 3 e 4, e que na quarta hora houve nova mudança significativa. Após vinte e quatro horas não se pode afirmar com certeza o que houve apenas com uma avaliação do gráfico dos *scores*, devendo, para maior um maior entendimento, ser avaliado o gráfico dos *loadings* e/ou histograma médio das freqüências de tons.

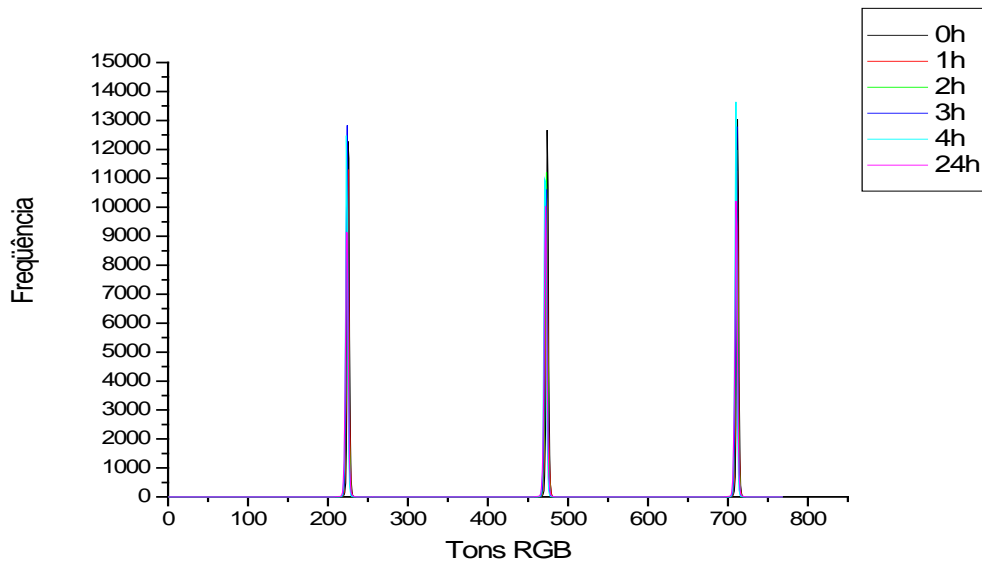
A PC1, Figura 19, representa quase toda a variância acumulada, pela PCA, sendo responsável por um percentual maior que 98%, sendo quase desprezível a variância acumulada pela PC2, em torno 0,3 %, o que favorece ainda mais a credibilidade do resultado apresentado.



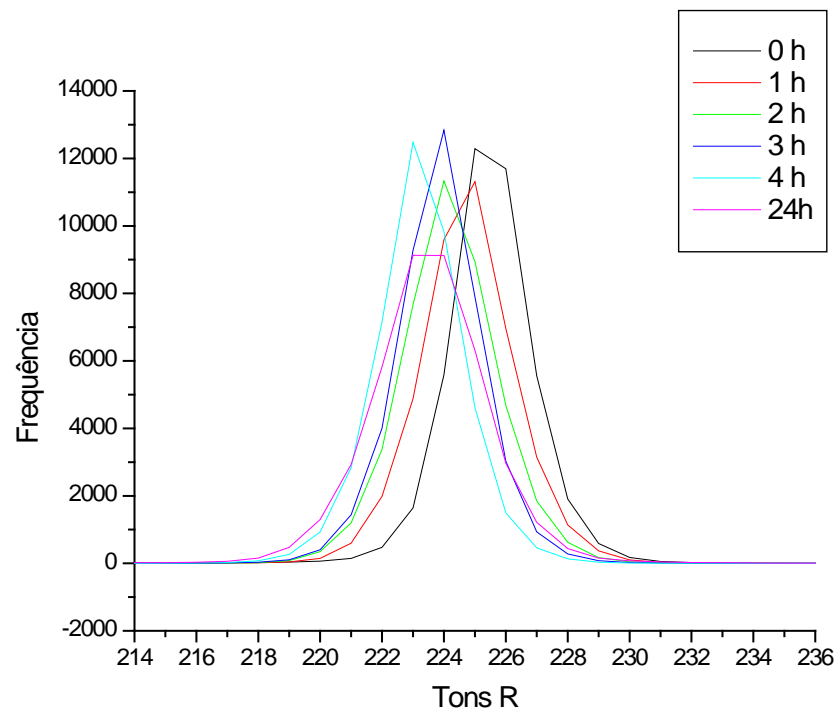
**Figura 20** - Imagens representativas da marca 1 em diferentes intervalos (0h, 1h, 2h, 3h, 4h e 24h) quando exposta à luz e ao ar. Imagens com resolução de 300 ppp no formato JPEG.

O histograma médio das freqüências de tons, possibilita verificar os intervalos de tons de cor que apresentam maiores número de repetições para um determinado índice, ou intervalo de índices, para cada canal do sistema RGB. Da Figura 21, que representa o histograma médio das freqüências para a marca 1 em diferentes intervalos de tempo é possível verificar que o intervalo dos índices de maiores freqüências são os seguintes: canal R, 216 a 232, conforme se verifica na figura 21.1, canal G, 464 a 480, figura 21.2 e canal B, 702 a 718, figura 21.3. Avaliando as figura 21.1, 21.2 e 21.3 é possível perceber que para o tempo 0 h há maior

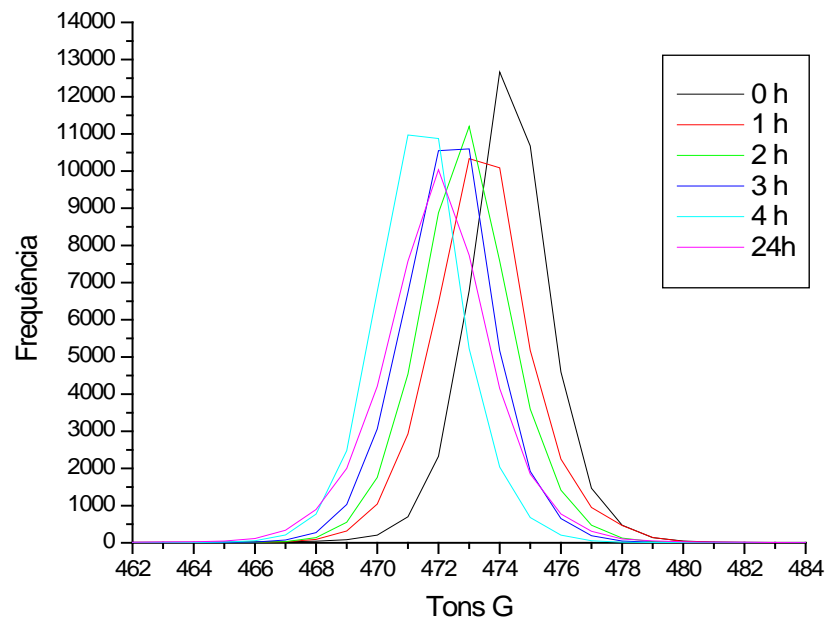
frequência para índices maiores, tons claros, enquanto que para o tempo 4 h para índices menores, tons escuros, podendo-se verificar que existe uma diminuição gradativa de frequência para índices menores a medida que se aumenta até o tempo de 4 h. Por sua vez no tempo de 24 h houve um comportamento diverso ficando com pico de frequência inferior aos demais e com a curva que não possibilita chegar-se a uma conclusão precisa do comportamento por meio do histograma médio de frequências.



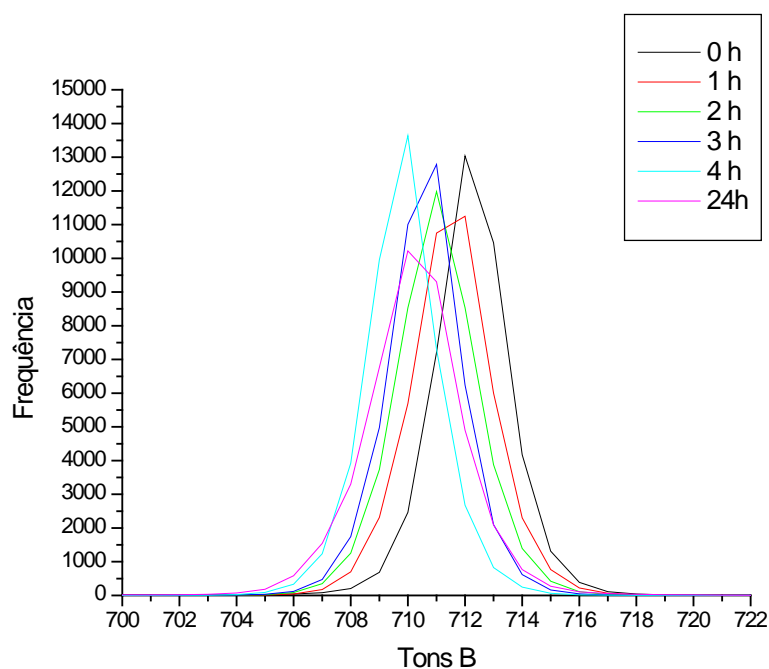
**Figura 21** – Histograma médio de frequências x Tons RGB de amostra da marca 1 em diferentes intervalos de tempo com resolução de 300 ppp no formato JPEG.



**Figura 21.1** – Histograma médio ampliado de frequências x Tons R de amostra da marca 1 em diferentes intervalos de tempo com resolução de 300 ppp no formato JPEG.



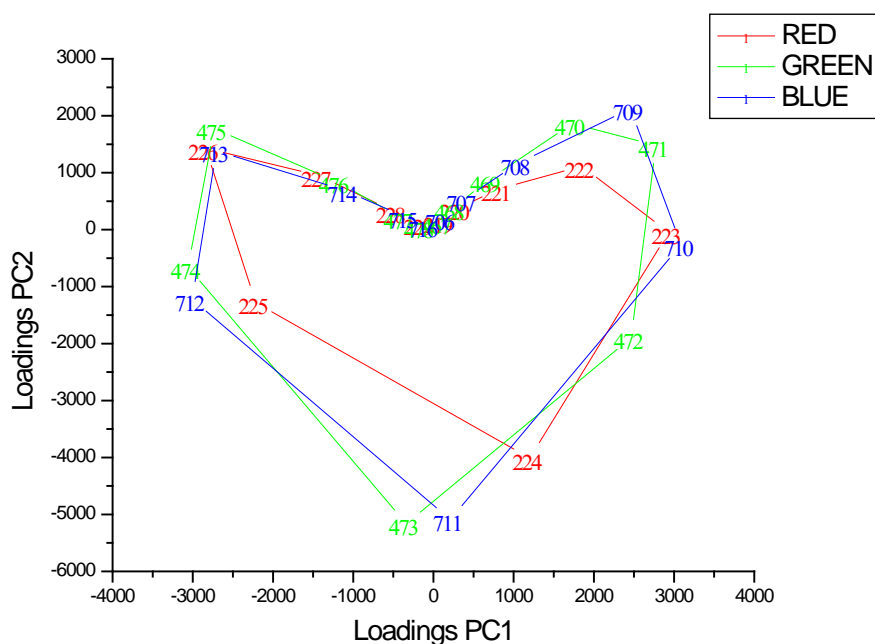
**Figura 21.2** – Histograma médio ampliado de frequências x Tons G de amostra da marca 1 em diferentes intervalos de tempo com resolução de 300 ppp no formato JPEG.



**Figura 21.3** – Histograma médio ampliado de freqüências x Tons B de amostra da marca 1 em diferentes intervalos de tempo com resolução de 300 ppp no formato JPEG.

O gráfico dos *loadings*, Figura 22, complementa o entendimento do gráfico dos *scores* e permite explicar o posicionamento do ponto 6, que refere-se à imagem obtida após vinte e quatro horas. No gráfico dos *loadings* os índices aumentam no sentido horário e pode ser observado para cada canal o seu comportamento. No canal R observa-se o aumento do índice 220 ao 228, que é o intervalo de índices com maior importância para a separação onde pode ser verificar que a região onde está o ponto 6, refere-se um valor próximo do índice 220 o que nos permite explicar como sendo o ponto 6 para uma imagem de cor mais escura. Os demais canais apresentam comportamento semelhante sendo possível perceber os seguintes índices como sendo os mais importantes, canal G: 469 a 477 e canal B: 707 a 715.





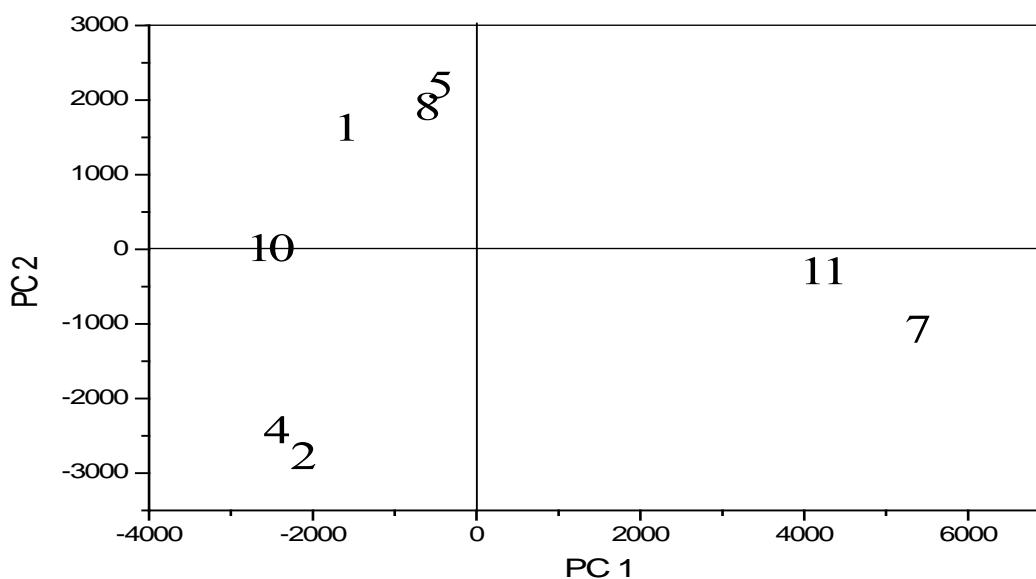
**Figura 22** - Gráfico dos *loadings* das duas primeiras componentes principais, PC1 e PC2 para a amostra da marca 1 com resolução de 300 ppp medidas no período de 0 h, 1 h, 2 h, 3 h, 4 h e 24 h. As numerações referem-se aos valores dos índices de cor para os três canais justapostos

#### 4.6 – Análise das imagens com resolução 300 ppp no formato BMP

Com o intuito de se ampliar o estudo de análise de imagem, utilizou-se de outro formato de arquivo de imagem. Para tanto, se fez uma nova amostragem de marcas de cervejas, onde foram escolhidas oito marcas, de acordo com a disponibilidade no mercado. Das oito marcas escolhidas, sete coincidiram com as marcas utilizadas para as análises anteriores, uma marca nova foi introduzida e três outras marcas que constavam dentre as amostras das análises anteriores não foram encontradas em mercado durante a nova amostragem. Desse modo passou a ter um novo conjunto de dados, onde a matriz de dados originais será uma matrix  $X_{8 \times 768}$ . Para o novo conjunto de dados foram executados os mesmos procedimentos para as análises com a matriz  $X_{10 \times 768}$ , mas com imagens capturadas no formato bitmap, BMP. O formato BMP foi escolhido por ser muito difundido e utilizado

para arquivos das imagens, tendo sido o precursor dos formatos de arquivos de imagem desde os idos do Sistema Operacional Windows 95®.

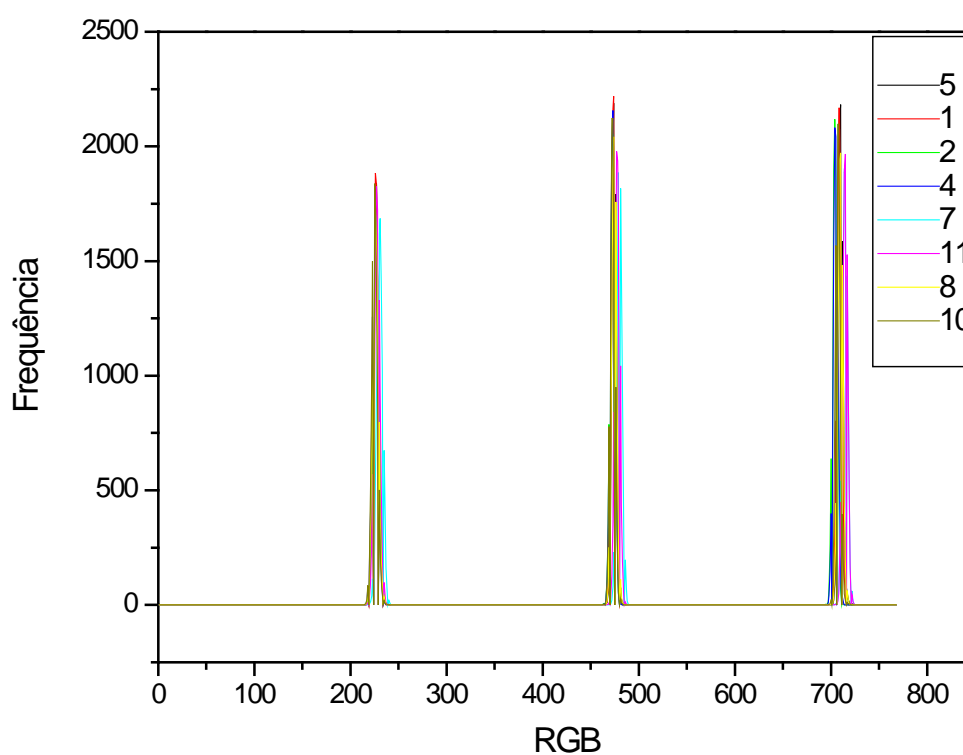
A Figura 23, apresenta o gráfico dos *scores* para as imagens de cervejas com resolução de 300 ppp no formato BMP. O gráfico bidimensional para as duas primeiras PCs tem variância acumulada maior que 90%, e a partir desse gráfico podemos observar uma boa separação das marcas de cervejas de onde pode-se verificar a formação de três agrupamentos e uma marca isolada. Em um agrupamento encontram-se as marcas 1, 8 e 5, onde as duas últimas devido à sua proximidade no gráfico pode-se afirmar que possuem grande similiaridade. Outro agrupamento é formado com as marcas 11 e 7, e outro agrupamento com as marcas 4 e 2, tendo ainda a marca isolada 10.



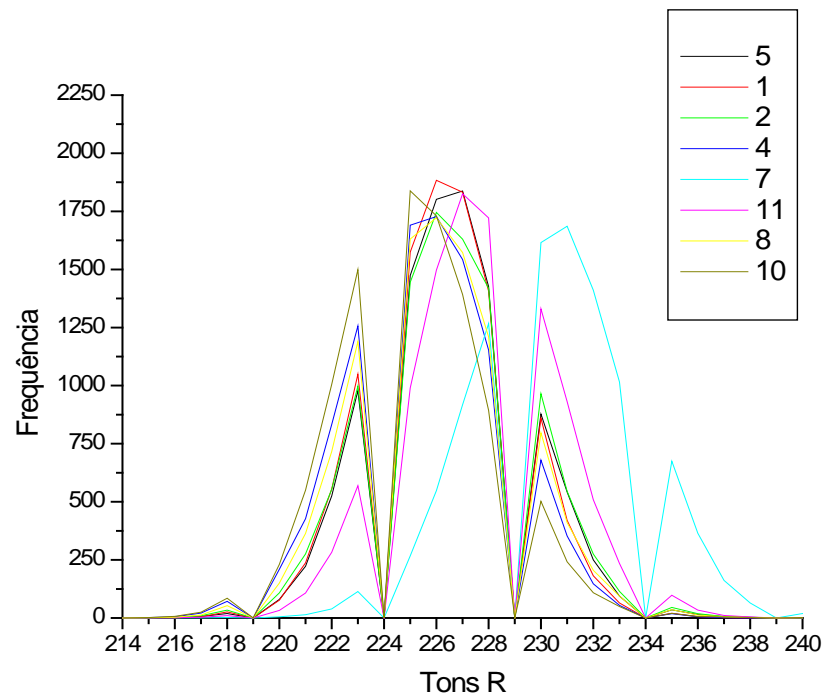
**Figura 23-** Gráfico dos *scores* PC1 e PC2 que explicam respectivamente 66,05 % e 25,50 % de variância dos canais RGB com 91,55% da variância acumulada dos dados das 8 diferentes marcas com resolução de 300 ppp no formato BMP

Convém observar que com a mudança da matriz de dados, com a retirada das marcas 3, 6 e 9 e a inserção da marca 11 a marca 2 até então tida como diferente das demais com relação a seu tom de cor, passou a ter considerável similaridade com a marca 4 nesse novo conjunto de dados.

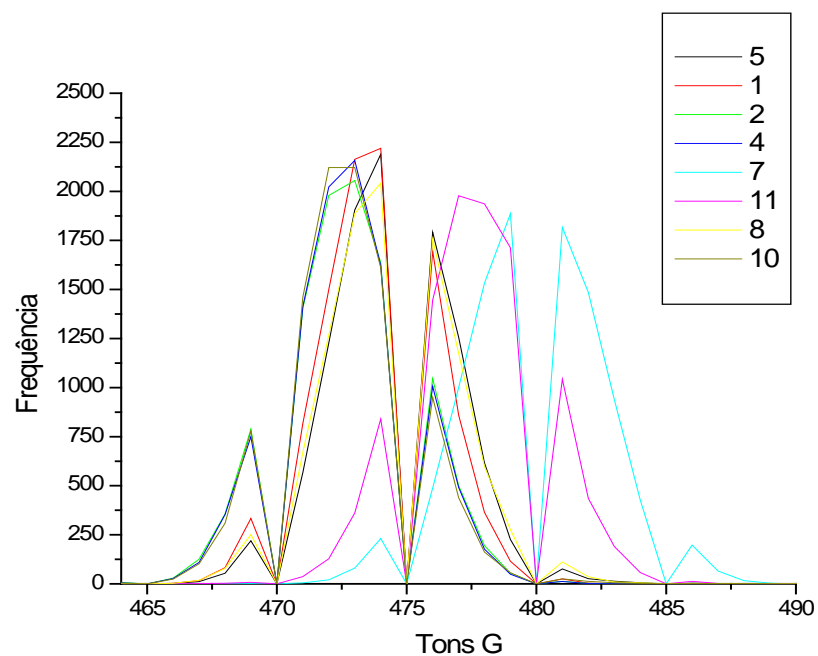
Cabe ressaltar que além da mudança da matriz de dados, mesmo com as marcas que se coincidem das análises anteriores, os lotes das marcas são diferentes, de modo que pode ter havido pequena diferenciação no produto final, fazendo com que a cor na mesma marca seja diferente por se tratar de lotes diversos.



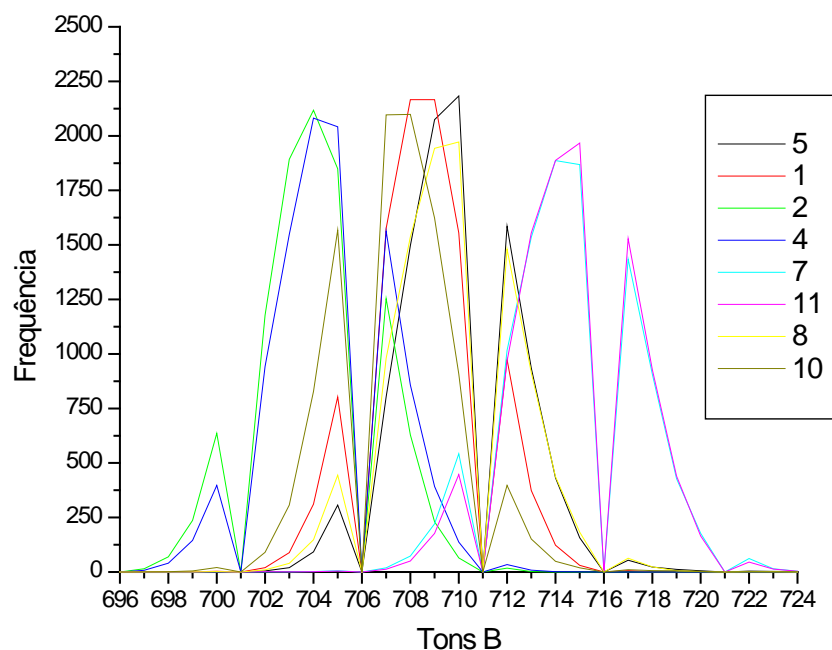
**Figura 24** – Histograma médio das frequências versus tons RGB de amostras de 8 marcas cervejas, com resolução de 300 ppp no formato BPM.



**Figura 24.1** – Histograma médio ampliado das frequências versus tons R de amostras de 8 marcas cervejas, com resolução de 300 ppp no formato BPM.



**Figura 24.2** – Histograma médio ampliado das frequências versus tons G de amostras de 8 marcas cervejas, com resolução de 300 ppp no formato BPM.

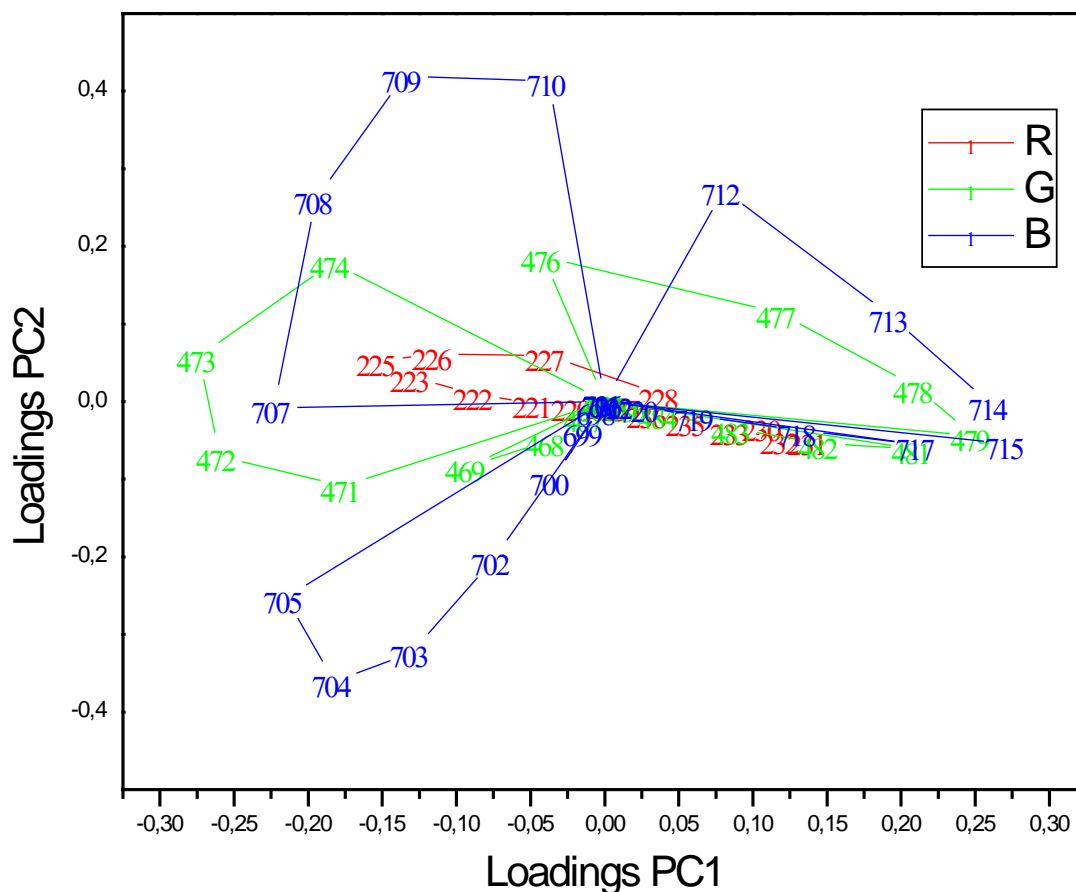


**Figura 24.3** – Histograma médio ampliado das freqüências versus tons B de amostras de 8 marcas cervejas, com resolução de 300 ppp no formato BPM.

A Figura 24, é o histograma médio das freqüências nos tons do sistema RGB, onde pode-se definir os respectivos intervalos para os índices do sistema R, G e B que possuem maiores freqüências possibilitando avaliar o comportamento de cada uma das marcas de cervejas. Os intervalos de maiores freqüências são: canal R, 216 a 240, figura 24.1, canal G, 465 a 490, figura 24.2, e canal B, 696 a 724, figura 24.3. Avaliando-se as figuras 24.1 e 24.2 é possível separar as marcas 7 e 11, as quais apresentam comportamento diferentes das demais e com maior freqüência para índices maiores, tons mais claros. Por sua vez a figura 24.3, específico para o canal B é possível separar além das marcas 7 e 11 com índices claros as marcas para tons escuros, menores índices, as marcas 2 e 4. Correlacionando o histograma do canal B com o gráfico de scores, figura 23, é possível confirmar as semelhanças destes dois pares. Avaliando-se os histogramas médios ampliados é possível afirmar que o canal B é o que permite a melhor classificação das marcas.

No gráfico dos *loadings*, Figura 25, é possível perceber uma disposição geométrica que é diferente das apresentadas pelos gráficos dos *loadings* nas análises anteriores. Comportamento diferente já havia sido percebido no histogramas médios das freqüências dos tons de cores para o canal B.

Pela disposição dos índices no gráfico dos *loadings*, observa-se que o canal B deve ser avaliado por três regiões diferentes e que contém os seguintes intervalos de índices: 699 a 705, 707 a 710 e 712 a 715, os quais estão aumentando em sentido horário. É possível perceber que os índices que faltam nos intervalos definidos são: 700, 706 e 711 os quais pelo gráfico têm peso zero ou próximo de zero, sendo índices de pouca importância para a classificação das marcas de cervejas. Esse fato explica o emaranhado de curvas nos histogramas médios das freqüências do canal B, onde pode-se perceber que para cada marca há mais de um pico, não tendo uma distribuição normal (tipo gaussiana) como observadas nos histogramas médios das freqüências das análises das imagens anteriores. Com as informações dos gráfico dos *scores* e de *loadings*, para a análise das marcas de cervejas no formato BMP com resolução de 300 ppp, as marcas referentes com tons de cores mais escuras são marca 2 e 4 e as marcas com tons de cores mais claras as marcas 11 e 7.



**Figura 25** – Gráfico dos *loadings* de PC1 e PC2 das marcas 8 de cervejas resolução de 600ppp, no formato BMP. As numerações referem-se aos valores dos índices de cor para os três canais justapostos.

No presente trabalho deve ficar claro que não foi priorizado qualquer tipo de marca de cerveja, as quais foram escolhidas ao acaso conforme disponibilidade local e os equipamentos utilizados eram os disponíveis no laboratório de Química Teórica e Computacional do Instituto de Química da Universidade Federal de Goiás.

## 5 - Conclusões

A análise das imagens possibilita a redução de dados multivariados das 768 variáveis do sistema RGB permitindo que essas imagens sejam avaliadas em duas dimensões, por meio de análise de componentes principais.

O uso do método alternativo proposto para a análise de cerveja com atributo de cor, por meio de PCA, mostrou ser possível classificar as cervejas por marcas, fazendo uso de equipamentos simples e acessíveis, o que evidencia o método como uma metodologia alternativa.

Foi possível avaliar a alteração de cor de cervejas expostas à luz e ao ar por meio de análise de imagens e PCA (gráfico de *scores* e *loadings*) e histograma médio de frequências dos índices RGB. As alterações de cor da cerveja exposta à luz e ao ar foram atribuídas aos fenômenos de oxidação das diversas substâncias químicas sujeitas ao escurecimento não enzimático.

As análises das imagens para amostras de cervejas com formato JPEG com resolução de 300 ppp e 600 ppp permitiram classificar as marcas de cerveja, sendo que cada uma das resoluções tiveram suas particularidades. A primeira (300 ppp) apresentou uma maior rapidez, com menores arquivos das imagens. A segunda (600 ppp) apresentou uma variância acumulada pouco maior, porém os arquivos das imagens são bem maiores. Assim, para cada análise deve-se fazer uma avaliação de custo/benefício e disponibilidade de recursos. Para o nosso trabalho a resolução em 300 ppp apresentou boa classificação com variância acumulada satisfatória, para as duas primeiras componentes principais com valores maiores que 70%.

As análises das imagens com formato BMP permitiram classificar as marcas de cervejas com variância acumulada alta, maior que 90%, porém por se tratar de um formato de imagem com dados não compactados, os arquivos gerados são grandes, de modo que com os equipamentos utilizados nesse trabalho as análises se tornaram demoradas.

Deve ficar claro que o método apresentado tem por objetivo separar



marcas, nada podendo ser afirmado além de suas pequenas diferenças de tons de cores, sendo até o presente momento o método visto apenas como complementar aos métodos já existentes.

Para classificação de marcas o método proposto é viável principalmente na atualidade com as facilidades computacionais para aquisição e manipulação das imagens e na obtenção de PCA, possibilitando que a utilização de métodos como esse se torne corriqueiro para a classificação, não só de cervejas, mas de qualquer gênero alimentício ou outro analito colorido.

Dessa forma pode-se concluir que o método proposto é uma excelente técnica auxiliar, para a classificação de cerveja, uma vez que não requer gastos e grandes treinamentos para a sua utilização, principalmente em um meio industrial, que ainda é considerado como restrito quanto aos seus métodos de análise e produção.

## ANEXO - A

**Tabela 1-** Algoritmo para extração de histogramas de frequências dos índices de cor dos canais do sistema RGB.

---

Algoritmo de obtenção de histograma de frequências.

---

```

imread('nome do arquivo de imagem.formato');
img_R=imhist(ans(800:1000,800:1000,1)); % índices de cor para o
canal R
img_G=imhist(ans(800:1000,800:1000,2)); % índices de cor para o
canal G
img_B=imhist(ans(800:1000,800:1000,3)); % índices de cor para o
canal B
Z1=[img_R' img_G' img_B'];% matriz contendo índices R, G e B da
imagem 1
imread('nome do arquivo de imagem.formato');
img_R=imhist(ans(800:1000,800:1000,1)); % índices de cor para o
canal R
img_G=imhist(ans(800:1000,800:1000,2)); % índices de cor para o
canal G
img_B=imhist(ans(800:1000,800:1000,3)); % índices de cor para o
canal B
Z2=[ img_R' img_G' img_B'];%matriz contendo índices R, G e B da
imagem 2

...

imread('nome do arquivo de imagem.formato');
img_R=imhist(ans(800:1000,800:1000,1)); % índices de cor para o
canal R
img_G=imhist(ans(800:1000,800:1000,2)); % índices de cor para o
canal G
img_B=imhist(ans(800:1000,800:1000,3)); % índices de cor para o
canal B
Zn=[ img_R' img_G' img_B'];%matriz contendo índices R, G e B da
imagem n

Wi=mean(Z1,Z2,...,Zn);%matriz média dos índices R, G e B para cada
marca

X=[W1;W2;...;Wi];% matriz contendo os índices R, G e B médios de
todas as marcas (matriz de dados originais)

```

---

## ANEXO – B

**Tabela 2-** Algoritmo para obtenção de PCA – matriz de loadings e scores e variância explicada por cada componente principal.

<b>Etapa 1-</b>	<b>Etapa 2-</b>	<b>Etapa 3-</b>
Pré-Processamento da matriz X (centrar na média)	Decomposição em Valores Singulares (SVD)	Variância explicada por cada Componente Principal
<pre>% 1. pré- processamento da matriz X [n,m]=size(X); Xm=mean(X); Xcm=X-ones(n,1)*Xm; % 1. centrar na média</pre>	<pre>% 2. decomposição em valores singulares [U S V]=svd(Xcm)  % 3. matriz de scores T=U*S</pre>	<pre>% 4. variância explicada por cada componente s=diag(S).^2; vart=sum(s); VARPC=(s/vart)*100</pre>

**ANEXO – C****Tabela 3-** Especificação das amostras das cervejas tipo pilsen digitalizadas nos formatos JPEG e BMP. A coluna “Local” representa a cidade de fabricação.

Categoria	Marca	Local	Teor Alcoólico (%)
Cerveja Pilsen	1	Brasília-DF	4,9
	2	Jacareí-SP	4,6
	3	Jacareí-SP	4,8
	4	Anápolis-GO	5,0
	5	Brasília-DF	4,8
	6	Alexânia-GO	4,5
	7	Trindade-GO	4,7
	8	Jacareí-SP	4,5
	9	Alexânia-GO	4,7
	10	Brasília-DF	4,7
	11	Alexânia-GO	4,7

**ANEXO – D****Tabela 4** – Especificação das marcas em resoluções 300 ppp e 600 ppp para o gráfico dos scores – Figura 15.

Amostra	Resolução em 300 ppp	Resolução em 600 ppp
Marca 1	1	11
Marca 2	2	12
Marca 3	3	13
Marca 4	4	14
Marca 5	5	15
Marca 6	6	16
Marca 7	7	17
Marca 8	8	18
Marca 9	9	19
Marca 10	10	20

## 6-REFERENCIAL TEÓRICO

ANTONELLI, M. COCCHI, P. FAVA, G. FOCA, G. C. FRANCHINI, D. MANZINI, A. ULRICI; *Automated evaluation of food colour by means of multivariate image analysis coupled to a wavelet-basic classification algorithm*, Anal. Chim. Acta, 515 3-13 2004.

APEXBRASIL. Associação Brasileira de Promoção de Exportações e Investimentos. Disponível em < [www.apexbrasil.com.br](http://www.apexbrasil.com.br) > Acesso em: 30 de setembro de 2005.

BORIN, A., FERRÃO, M. F., MELLO, C., CORDI, L., PATACA, L. C. M., DURÁN, N., and POPPI, R. J.; *Quantification of Lactobacillus in fermented milk by multivariate image analysis with least-squares support-vector machines*. Anal. Bioanal. Chem., 387, 1105–1112 (2007).

BRUNS, R.E. E FAIGLE, J. F. G.; *Quimiometria*; Química Nova, p 84-99, Abril de 1985.

CERVESIA. Disponível em <http://www.cervesia.com.br>. Acesso em: 18/09/2006.

CORREIA, P. R. M.; FERREIRA, M. M. C., *Reconhecimento de padrões por métodos não supervisionados: explorando procedimentos quimiometricos para tratamento de dados analíticos*, Química Nova, Vol. 30, p 481-487, 2007.

Copyright © 1989-2005. SIP ENPC. Scilab é uma marca da SIP.

DESCHAMPS, F; *Contribuições para o desenvolvimento de um sistema de visão aplicada ao monitoramento do desgaste de ferramentas de corte – O Sistema Toolspy*; 2004. 158f; Dissertação de Mestrado, UFSC, 2004.

DECRETO nº 2.314 de 04 de setembro de 1997. Regulamenta a Lei nº 8.918, de 14 de julho de 1994, que dispõe sobre a padronização, a classificação, o registro, a inspeção, a produção e a fiscalização de bebidas. In: Presidência da República, Casa Civil, Subchefia para Assuntos Jurídicos. Brasília, 1997. Disponível em <[https://www.planalto.gov.br/ccivil\\_03/decreto/d2314.htm](https://www.planalto.gov.br/ccivil_03/decreto/d2314.htm)> Acesso em: 30 abr. 2008.

ESBENSEN, K. e GELADI, P., *Strategy of multivariate image analysis (MIA)*, Chemometrics and Intel. Lab. Systems, 7, 67-86-1989.

FERREIRA, M.M. C.; ANTUNES, A. M.; MELGO, M. S. ; VOLPE, P. L. O.; *Quimiometria I: calibração multivariada, um tutorial*; Química Nova, vol 22, p. 724 – 731, 1999.

GELADI, P; BENGTSSON, E.; ESBENSEN, K.; GRAHN, H., *Imagen analysis in chemistry I. Properties of image analysis*, Trends in Analytical Chemistry, 11, 121-131, 1992.

GELADI, P. e GRAHN, H.N; *Multivariate image analysis*, John Wiley & Sons, England, 1996.

GELADI, ISAKSSON H, LINDQSVIST, L, WOLD, S, ESBENSEN, K; *Principal Component Analysis of Multivariate Images*, Chemom, Intell. Lab. Syst, 5, 209-220, 1989.

GELADI, P.; KOWALSKI, B. R. *Partial Least-Squares Regression: A tutorial*. Anal. Chim. Acta, v. 185, n.1 , p. 1-17, 1986.

HAYAKAWA, T.; HIRAI, M.; *An assay of ganglioside using fluorescence image analysis on a thin-layer chromatography plate*. Anal. Chem. 75, 6728-6731, 2003.

HIRANO, A.; MORIDERA, N.; AKASHI, M.; SAITO, M.; SUGAWARA, M.; *Imaging of l-glutamate fluxes in mouse brain slices based on an enzyme-based membrane combined with a difference-image analysis*. Anal. Chem, 75, 3775-3783, 2003.

LAQQA, disponível em: <<http://laqqa.iqm.unicamp.br/>> Acesso em: 20/09/2007.

Lei 8.918 de 14 de julho de 1994. Dispõe sobre a padronização, a classificação, o registro, a inspeção, a produção e a fiscalização de bebidas, autoriza a criação da Comissão Intersetorial de Bebidas e dá outras providências. In: Presidência da República, Casa Civil, Subchefia para Assuntos Jurídicos. Brasília, 1994. Disponível em <[https://www.planalto.gov.br/ccivil\\_03/leis/l8918.htm](https://www.planalto.gov.br/ccivil_03/leis/l8918.htm)> Acesso em: 30 abr. 2008.

LYNCH, M. e WOOLGAR, S. (editores), *Representation in Scientific Practice*, The MIT Press, London, 1990.

MATOS, G. D.; FILHO, E. R. P.; POPPI, R. J.; ARRUDA, M. A. Z., *Análise exploratória em Química Analítica com emprego de Quimiometria: PCA e PCA de imagens*, Revista Analytica, p 38-50, 2003.

MATTHIAS R. REINOLD; *Manual Prático de Cervejaria*. Aden Editora, São Paulo, 1997.

MELLO, C. A.; *Redes neurais e técnicas de pré-processamento de sinais em espectroscopia de reflectância no infravermelho próximo*; 1998. 147 f.; Tese (doutorado), Instituto de Química, Universidade Estadual de Campinas. 1998

MESSERSCHMIDT, I.; *Análise quantitativa por espectroscopia no infravermelho médio empregando técnicas de reflectância e calibração multivariada*; 138 f.; Tese (doutorado), Instituto de Química, Universidade Estadual de Campinas, Campinas, 1999.

MINGOTI, A. S.; *Análise de dados através de métodos de estatística multivariada: Uma abordagem aplicada*. Editora UFMG, Belo Horizonte, 2005.

MORGANO, M. A.; QUEIROZ, S. C. N.; FERREIRA, M. M. C., *Determinação dos teores de minerais em sucos de frutas por espectrometria de emissão óptica em plasma indutivamente acoplado (ICP-OES)*, Ciência e Tecnologia de Alimentos, vol. 19. p. 344-348, 1999.

MOURA, E. R.; *Determinação do fármaco maleato de enalapril em medicamento por espectroscopia de reflectância difusa na região do infravermelho médio associado à calibração multivariada*, Dissertação (mestrado); Instituto de Química, Universidade Federal de Goiás, p 32-37. 2005.

NETO, J. M. M.; MOITA, G. C.; *Uma Introdução à análise exploratória de dados multivariados*, Química Nova, p 67-69, vol. 21. 1998.

NOBLE, B.; DANIEL, J.W.; *Álgebra linear aplicada*; Rio de Janeiro: Prentice-Hall do Brasil, 1986.

PATACA, Luiz C. M. ; BORGES NETO, W. ; MARCUCCI, M. C. ; POPPI, R. J.; *Determinação de cor de mel de abelhas utilizando análise multivariada de imagens*. In: 28 Reunião Anual da Sociedade Brasileira de Química, 2005, Poços de Caldas - MG. 28 Reunião Anual da Sociedade Brasileira de Química - Química para o Desenvolvimento Sustentável e Inclusão Social. São Paulo: Copypress, 2005. v. 2005.

PRATT, W.K.; *Digital Image Processing*, 2nd Ed., John Wiley & Sons: New York, 1991.

REICHENBACH, S. E.; NI, M.; KOTTAPALLI, V.; VISVANATHAN, A.; *Comparative visualization for comprehensive two-dimensional gas chromatography*, *Chemometr. Intell. Lab. Syst.* 71-107, 2004.

SCHIMIDT, F., *Desenvolvimento de um programa computacional para o tratamento de imagens digitalizadas e sua aplicação em Química Analítica*, Dissertação (mestrado), Instituto de Química - Unicamp, Campinas, 1997.



SILVA, J. B. P.; MALVESTITI, I.; HALLWASS, F.; RAMOS, M. N.; LEITE, L. C.; BARREIRO, E. J.; *Aplicação de análise de componentes principais para verificação de atribuições de sinais nos espectros de RMN H. O caso dos 3-aril (1, 2, 4)-oxadiazol-5-carboidrazida benzilidenos*, Química Nova, p. 492-496 vol. 28, 2005.

SIP – *Scilab Image Processing Toolbox*. Copyright © 2002-2004, Ricardo Fabri. USP, São Paulo, 2003.

SIP stands for **S**cilab **I**mage **P**rocessing toolbox, Disponível em <http://siptoolbox.sourceforge.net/> Acesso em: 15 mar. 2007.

SOUZA, L. A.; SPEHAR, C. R.; SANTOS, R. L. B., *Análise de imagem para determinação do teor de saponina em quinoa*, Pesquisa Agropecuária Brasileira, vol. 39, p 397-401, 2004.

YU H. L. J.F.MACGREGOR; *Multivariate image analysis and regression for prediction of coating content and distribution in the production of snack foods*. Chemo. Intell. Lab. Sys. 67 (2), p. 125-144, Aug 28, 2003.