Instituto de Química
Programa de Pós-graduação em Química

DISSERTAÇÃO DE MESTRADO

IDENTIFICAÇÃO DE FRAUDES EM DOCUMENTOS POR ADIÇÃO DE TEXTO E OBLITERAÇÃO UTILIZANDO IMAGENS HIPERESPECTRAIS E QUIMIOMETRIA

ALUNA: THAYNA APARECIDA DE OLIVEIRA

ORIENTADOR: PROF. DR. JEZ WILLIAN BATISTA BRAGA

COORIENTADOR: DR. MÁRCIO TALHAVINI

Brasília, DF
2019

Universidade de Brasília
Instituto de Química (IQ)

FOLHA DE APROVAÇÃO

Comunicamos a aprovação da Defesa de Dissertação do (a) aluno (a) **Thayna Aparecida de Oliveira**, matrícula nº **17/0089118**, intitulada "***Identificação de fraudes em documentos por adição de texto e obliteração utilizando imagens hiperespectrais e quimiometria***", apresentada no (a) Auditório Azul do Instituto de Química (IQ) da Universidade de Brasília (UnB) em 31 de julho de 2019.

Prof. Dr. Jez Willian Batista Braga
Presidente de Banca

Prof. Dr. Alexandre Fonseca
Membro Titular

Prof. Dr. Fábio Moreira da Silva
Membro Titular IQ/UnB

Dr. Jorge Jardim Zacca
Membro Suplente

Em 31 de julho de 2019.

AGRADECIMENTOS

Ao Deus que criou todo este universo maravilhoso. Que fez os céus, a Terra, o mar e tudo que neles há. Que me fez e teceu cada um dos dias da minha vida quando nenhum deles ainda existia. A esse Deus que é Excelso, Poderoso, mais do que eu jamais vou conseguir mensurar na minha mente finita e limitada. A esse Deus que se fez homem, se fez carne e se fez homem para me dar a vida eterna por meio do meu Senhor Jesus Cristo. Ao Deus que me tirou da morte para a vida por meio do seu único Filho e que por Ele me deu acesso à vida eterna. A esse Deus que me tornou filha amada e cuidada. A esse Deus que é o Senhor Soberano de todo universo criado e que é o meu Pai que enxuga as minhas lágrimas e dá a alegria e a beleza da vida debaixo do Sol. Ao Deus triúno que consola meu coração com a alegria do Espírito Santo e que traz esperança quando eu não a tenho. Ele é perfeitamente aquilo que eu não sou. Ao amor que é a razão que dá sentido à beleza da vida daqui, em dias pesados, em dias difíceis, em dias incríveis, em dias comuns. É tudo por Ele, para Ele e sobre Ele. E isso é maravilhoso. Viver à luz da redenção sabendo que todo dom, conhecimento, sabedoria e entendimento vêm do Senhor e que nEle não há variação nem sombra de mudança. Ao Deus que tem todos os tesouros da sabedoria e do conhecimento escondidos em Cristo. A Ele que me chama a amá-lo com todo o meu coração, alma, força e entendimento. A Cristo que é a razão de tudo e é a razão que há em mim. Obrigada, meu Senhor, por ser o Soberano de todo o universo e por ser o meu Pai que enxugou tantas lágrimas durante a caminhada até aqui e que também deu a razão de todos os meus sorrisos.

À minha amada família que o meu Senhor bondosamente me concedeu. Papai, mamãe e The, vocês são muito mais do que eu mereço e são verdadeiramente uma terna expressão do amor e da bondade do Senhor para comigo: obrigada por orarem por mim e comigo, obrigada pela paciência em me ouvirem, em enxugarem as minhas lágrimas, em trazerem palavras de admoestação e disciplina, em trazerem palavras de conforto e de esperança, em ficarem em silêncio e apenas me ouvirem, por chorarem comigo e por se alegrarem. A vida é mais bonita porque vocês estão nela e me lembram, um dia após o outro, que o meu Deus é bom para comigo. Amar vocês é graça do Senhor para comigo e ser amada por vocês é experimentar todos os dias a bondade e a graça abundante do Senhor na minha vida. Obrigada por insistirem tantas vezes e por me amarem conhecendo tão intimamente minhas falhas e permanecerem mesmo assim. Eu os amo e sou grata ao Senhor por Ele ter escolhido vocês para serem o meu papai, a minha mamãe e o meu irmão, tão amados, tão chegados, tão queridos. Perdoem-me por todas as minhas ausências e falhas quando não pude desfrutar de mais tempo com vocês.

À minha amada família da fé que o Senhor me deu: amados pastores, irmãs e irmãos tão queridos, eu sou grata ao nosso Senhor por ter irmãos e irmãs que me sustentam em oração, que dividem o fardo comigo e que se alegram comigo. Obrigada por tantas vezes terem me escutado dizer que eu estava tão cansada e que eu apenas queria desistir e por terem me trazido, tão reiteradamente, palavras de esperança ao meu coração e me lembrarem que o nosso Senhor tem o tempo certo de todas as coisas, está absolutamente no controle de tudo e faz com que todas as coisas cooperem para formar o caráter do

nosso Senhor em nós, mesmo em dias que parecem tão escuros. Muito obrigada por cada oração tão tarde da noite, por cada palavra de compreensão, de bom ânimo e de repreensão que vocês me deram. Quanta sabedoria em conselhos tão pacientes e tão insistentes para comigo. Como sou grata ao Senhor por vocês. Eu os amo e caminhar aqui com vocês é graça do Senhor na minha vida.

Aos meus ternos amigos: vocês conhecem a minha alma e eu não consigo mensurar em palavras a bondade do meu Senhor por ter me concedido vocês. Obrigada por chorarem comigo e por se alegrarem. Por me ouvirem, por tantos meios diferentes e em situações tão diversas. Por cada café, sobremesa, almoço, chocolate ou tempo desfrutado juntos. Por terem sido tão gentis expressão do cuidado do meu Senhor para comigo. Vocês são raros e tão preciosos para mim. Espero que o nosso Senhor permita sempre que a nossa amizade cresça sempre, firmada na razão da nossa esperança: Cristo. Vocês foram ânimo e força quando eu não as tinha mais, por meio de palavras, orações, abraços, ouvidos. Por meio de uma amizade tão valorosa que aqueceu meu coração. Eu os amo e ter vocês na vida debaixo do Sol é bondade do Senhor para comigo.

Aos meus colegas de trabalho e a toda a equipe que caminhou comigo durante esse tempo: obrigada pela paciência, pela disponibilidade, pelo tempo concedido e por me escutarem tantas e tantas vezes sobre um assunto tão diferente mas que me faz brilhar os olhos: química é algo tão bonito. E eu acredito que no fim vocês também conseguiram enxergar um pouquinho da beleza da química que está por todos os lados.

Aos meus queridos orientador e coorientador: obrigada pela paciência e pela humildade em ensinar e demonstrar a beleza que existe na química e na quimiometria. É muito bom ver pessoas exercerem com talento aquilo que elas fazem. E vocês o fazem muito bem. Sou grata a Deus por Ele ter me concedido a alegria de ser orientada por vocês. Foi um prazer e um grande aprendizado para mim.

À toda a Universidade de Brasília e à equipe da Polícia Federal: obrigada pelas portas abertas para compartilharem e construírem conhecimento, com humildade, disponibilidade e alegria.

Cl 1: 13-20

Sl 115:1

Sl 96:4-6, 9

Gl 2:20

Sl 111:10

Pv 1:7

RESUMO

A análise de documentos visando a identificação de alteração feitas por adição de partes de números ou obliteração de trechos de texto é uma área de crescente interesse na ciência forense e tem demandado o desenvolvimento de técnicas que associem características típicas à análise de amostras forenses, como a preservação da integridade da amostra, com o requisito de objetividade na análise. A combinação de técnicas espectroscópicas, como a análise hiperespectral, e de quimiometria, possibilita a união desses atributos, fornecendo informações acerca da composição química e da distribuição dos componentes na amostra. O Comparador Espectral de Vídeo - VSC6000®, aparelho comumente encontrado nos laboratórios das polícias do Brasil, possui a função hiperespectral e, com o uso de ferramentas quimiométricas, é possível ampliar o uso desse equipamento nas análises forenses. O objetivo dessa pesquisa foi associar o uso da função imagem hiperespectral do VSC6000® à análise multivariada, e propor o desenvolvimento de um método não destrutivo, rápido e objetivo para identificação de fraudes em documentos por adição de texto e por obliteração. Foram testadas amostras de canetas azuis esferográficas, de 17 marcas e modelos distintos, que não são distinguíveis visualmente, com textos produzidos em papel sulfite, com análise na faixa espectral de 400 a 1000 nm. Para a adição de texto, as amostras simularam a adição de um algarismo "zero" a um número dez, produzindo uma falsificação do número cem. Para a obliteração, as amostras simularam a revelação de quantidade de letras da palavra "FALSO", escrita por uma caneta e ocultada pela sobrecarga de outra. Demonstrou-se que, pela combinação de métodos de análise univariada e multivariada - utilizando-se Análise de Componentes Principais e Resolução Multivariada de Curvas - é possível a identificação de fraudes por adição de texto em 92% dos casos analisados, e a revelação de toda ou parte da informação escondida para 93% dos casos de obliteração analisados. Nos casos da fraude por obliteração, a realização de um teste cego comprovou a eficácia de 67% de acerto na identificação da fraude. Demonstrou-se que a associação da imagem hiperespectral à quimiometria, em casos de adição de texto e obliteração, é capaz de potencializar uma técnica baseada em um aparelho com baixa resolução espectral, como o VSC6000®, e aumentar a sua eficácia para a solução de casos nos laboratórios de polícia brasileiros.

Palavras-chave: Imagem hiperespectral, Documentoscopia, Forense, VSC, PCA, MCR.

ABSTRACT

The analysis of document alteration through the addition of parts of numbers or obliteration of text segments is an area of growing interest in forensic science and has required the development of techniques that associate typical characteristics to the analysis of forensic samples, such as preservation of the integrity of the sample, with the requirement of objectivity in the analysis. The combination of spectroscopic techniques and multivariate statistics allows the union of these attributes. The association of hyperspectral imaging with chemometrics provides information about the chemical composition and distribution of components in the sample. The purpose of this work is the development of a non - destructive, fast and objective method for identification of document fraud by addition of text and hyperspectral image - based obliteration with the use of the Video Spectral Comparator - VSC6000 ® and chemometrics. Samples of blue ballpoint pens of 17 different brands and models, which are not visually distinguishable, with texts produced in sulphite paper, with analysis by the spectrometric function in the 400 to 1000 nm range, were tested. The samples simulated: for the addition of text, the addition of a digit "zero" to a number ten, producing a falsification of the number one hundred; for obliteration, the revelation of the quantity of letters of the word “FALSE”, written by one pen and concealed by the overload of another. It has been demonstrated that by combining univariate and multivariate analysis methods - using Principal Component Analysis and Multivariate Curve Resolution - it is possible to identify fraud by adding text in 92% of the analyzed cases, and the disclosure of part or total hidden information to 93% of the cases of obliteration analyzed. It was verified that, in the cases of obliteration fraud, there is variation of the result due to the previous knowledge of the hidden word. Multivariate analysis with the use of hyperspectral imaging data has been reported in the literature as an appropriate tool for the discrimination of pen paints and it has been demonstrated, through the present study, that chemometrics is capable of potentiating a technique based on a spectral low-resolution device, such as VSC6000®, and increase its effectiveness for case resolution in Brazilian police laboratories.

Keywords: Hyperspectral image. Documentoscopy. Forensic. VSC. PCA. MCR.

ÍNDICE

LISTA DE ABREVIATURAS E ACRÔNIMOS

HSI - Imagem Hiperespectral (do inglês, *Hyperspectral Image*).

PCA - Análise por componentes Principais (do inglês, *Principal Component Analysis*).

VSC® - Comparador Espectral de Vídeo (do inglês, *Video Spectral Comparator*)

RGB - Vermelho, Verde e Azul (do inglês, *Red, Green and Blue*).

PC - Componente Principal (do inglês, *Principal Component*).

SVD - Decomposição por Valores Singulares (do inglês, *Singular Value Decomposition).*

MCR - Resolução Multivariada de Curvas (do inglês, *Multivariate Curve Resolution*).

IR – Infravermelho (do inglês, *Infrared*)

UV – Ultravioleta

LISTA DE FIGURAS E TABELAS

1. Introdução e objetivos

1.1. Introdução

A ciência forense pode ser entendida como a ciência que auxilia a justiça na definição do que é verdade, envolvendo a aplicação de conhecimento científico, técnico ou especializado na resolução de questões cíveis ou criminais[1]. A química forense é o ramo da ciência forense voltado à produção de provas materiais para a justiça, através da análise química em diversas matrizes, como drogas lícitas e ilícitas, venenos, acelerantes, resíduos de incêndio, explosivos, resíduos de disparo de armas de fogo, combustíveis, tintas e documentos[2]. A Documentoscopia é definida como a área da criminalística que estuda os documentos a fim de verificar sua autenticidade e apontar a sua autoria[3] e as alterações documentais têm se destacado como uma área de importância crescente devido à quantidade de ocorrências na área pericial.

A análise de tintas, de papéis e de suas interações tem sido uma importante área de estudo em ciência forense. Seu principal objetivo é verificar as adulterações em documentos[4]. Quando a caneta é o material de escrita dos documentos questionados, as evidências da existência de duas ou mais tintas no documento podem ser informações valiosas, indicando que o documento foi modificado pela adição ou substituição de elementos[5].

A fraude por meio de alteração e adição de partes de números ou de texto é um tipo recorrente de caso encontrado em exames forenses de documentos[6,7]. Crimes envolvendo a obliteração de partes da escrita também são muito comuns. Nesses casos, o objetivo é possibilitar a leitura das informações que foram cobertas por uma tinta sobreposta[7].

Em uma análise forense, pela necessidade legal típica para a prova material, a integridade da evidência documentoscópica coletada deve ser preservada sempre que possível. Portanto, uma análise destrutiva deve ser realizada somente se um método não destrutivo não estiver disponível para resolver o caso[4].

Um dos equipamentos mais empregados nas análises periciais na área de documentoscopia é o Comparador Espectral de Vídeo (VSC®, do inglês *Video Spectral Comparator*)[4]. O VSC® é um instrumento para exames que exigem a detecção de

diferenças nas propriedades ópticas e na composição química das tintas de canetas. O aparelho combina várias fontes luminosas (como luz visível (VIS), ultravioleta (UV) e infravermelho (IR)) e filtros com a qualidade de imagem digital. É um aparelho de simples operação e que permite uma análise não-destrutiva – características que o tornam uma ferramenta muito eficaz na resolução de vários casos investigados no dia-a-dia das polícias brasileiras[4,5,6]. Contudo, apesar dessa versatilidade, as análises espectrométricas realizadas pelo VSC® não são tão precisas como as de um espectrômetro[18] e o comparador espectral de vídeo tem se mostrado eficaz apenas em situações nas quais é possível encontrar um único comprimento de onda seletivo que diferencie as tintas envolvidas na análise. A análise multivariada, por meio da interpretação estatística dos dados, permite ampliar o uso do aparelho na realização dos casos periciais.

Técnicas analíticas não-destrutivas e rápidas são as ideais para análises forenses e tal combinação justifica o crescente uso dessas na documentoscopia[8]. As técnicas espectroscópicas que preservam a integridade da amostra na análise são capazes de reunir esses dois atributos e isso justifica o crescente uso dessas na documentoscopia.

Com o objetivo de tornar a análise dos dados espectroscópicos mais efetiva, os pesquisadores têm utilizado a análise estatística avançada, por meio da qual é possível obter estimativas quantitativas em análises conclusivas e tal característica é muito útil para aplicações forenses[9]. São obtidos melhores resultados quando as técnicas analíticas são computáveis e correlacionadas com a confiança estatística, de modo que a possibilidade de erros é minimizada[10]. Nesse sentido, o desenvolvimento de métodos analíticos combinados com ferramentas quimiométricas tem se mostrado crescente em várias áreas por proporcionar uma análise de dados mais eficaz[11]. A quimiometria é uma área da Química, extremamente, difundida nos dias atuais e muito útil na extração de informações dos mais variados sistemas químicos[12].

No exame de documentos questionados, por meio da análise de tintas presentes em cores ou em composições diferentes, podem ser utilizados métodos simples de imagens, não-destrutivos e sem preparo de amostra. Entretanto, a inspeção visual ou a utilização de imagens em escalas (RGB, do inglês *Red, Green, Blue*) podem ser insuficientes quando as propriedades de cor das tintas a serem discriminadas forem muito semelhantes entre si. Nesses casos, pode-se valer dos sistemas de imagens

hiperespectrais[7]. A imagem hiperespectral combina imagens convencionais e espectroscopia para obter informações espaciais e espectrais (químicas) de uma amostra[13,7]. Assim, a identidade e a distribuição de cada componente químico ao longo da imagem é revelada[7].

O uso de imagens hiperespectrais em análise multivariada tem demonstrando elevada potencialidade para análise de falsificações documentais por adição de texto e por obliteração[14,1,7], mas ainda pouco explorada com o uso de instrumentos disponíveis nos laboratórios forenses, como o VSC®. É essa abordagem que o presente estudo se propõe a realizar.

1.2. Objetivo Geral

O objetivo principal desta pesquisa é a avaliação do uso de ferramentas quimiométrica para ampliar o uso do aparelho Comparador Espectral de Vídeo - VSC6000®, empregando o modo de aquisição hiperespectral, para a identificação de fraudes em documentos por adição de texto e obliteração quando a simples análise por seleção de comprimento de onda não demonstrar eficácia para a solução desses tipos de casos em aplicações forenses.

1.3. Objetivos específicos

- Propor um método simples e que amplie o uso de um equipamento de uso frequente nos laboratórios de análise de perícia documentoscópica disponíveis nas Polícias Federal e Civis a fim de gerar mais eficiência e eficácia para as análises periciais.

- Propor um método objetivo e com protocolo simples, com o uso de uma técnica não – destrutiva, baseada em imagem hiperespectral e análise multivariada, para identificar fraudes por adição de texto e por obliteração em documentos produzidos com canetas esferográficas de cor azul, de diversas marcas, disponíveis no comércio brasileiro.

- Aplicar o método na identificação de fraudes por adição de texto e por obliteração feitas com canetas de tintas de cor azul em documentos cujo suporte seja papel sulfite.

- Avaliar a eficiência de um modelo semi-quantititativo de análise multivariada para a identificação de fraudes por obliteração.

2. Revisão Bibliográfica

2.1. Métodos de análise de tintas em documentos

Um documento impresso pode ser definido como todo suporte material (papel, plástico, dentre outros) o qual contenha informação útil (impressões, manuscritos, imagens). Documentos fazem parte do cotidiano da vida das pessoas e são utilizados em diversas aplicações, como celebração de contratos, comprovação de filiação, dentre outros usos. Por isso, os documentos são objeto comum de fraudes por parte dos criminosos [16]. No caso da Polícia Federal, segundo dados do documento "Relatório de Gestão Consolidado", referente ao exercício de 2016, aproximadamente 19,87% dos laudos produzidos pelos peritos criminais federais naquele ano são relativos à análise pericial de documentos[15].

A documentoscopia tem como objetivo, por meio de exames e de comparações, determinar a autenticidade e a autoria de documentos[16]. A perícia documentoscópica busca obter informações como a identidade do escritor, o conteúdo de textos apagados, dentre outras, e utiliza diversas estratégias para cumprir seus objetivos. Dentre esses, estão os relacionados à verificação de alteração documental em que se busca investigar a existência de alguma alteração física no documento que modifique sua essência ou o seu teor original. Tais alterações podem envolver rasuras, obliterações, acréscimos, dentre outros tipos de fraudes[1].

A busca por técnicas não-destrutivas, menos subjetivas e mais eficientes (com um mínimo de falsos positivos ou negativos) tem sido uma realidade crescente na química forense, inclusive no que tange à Documentoscopia. Pereira et al[6]. citam, como requisitos essenciais a uma técnica de análise forense documental, a preservação da integridade dos documentos; a objetividade das informações obtidas e a velocidade de análise.

A análise de tintas é um dos tipos de exames que pode ser realizado para verificar alterações documentais. Esse estudo é realizado para verificar a ocorrência de falsificações[4] e compreende análise de tintas em documentos, com a finalidade de determinar sua composição ou suas características físicas e/ou químicas para sua identificação ou diferenciação[1]. A análise documental envolve o exame da tinta da

caneta para investigar se a mesma foi utilizada em duas ou mais versões de textos manuscritos relacionados a casos de mudanças ou de adições de texto a um documento[4].

2.2. Principais métodos para análise de tintas

As tintas de caneta são compostas por uma diversidade de componentes como colorantes (corantes ou pigmentos), solventes, resinas, surfactantes, lubrificantes, emulsificantes e aditivos. Todos esses componentes variam em suas composições e combinações, de produto a produto. E todas essas diferenças são vitais para a análise forense a fim de determinar quantas tintas foram utilizadas no documento, se mais de uma tinta for encontrada nele. A análise é feita pela comparação entre amostras de tintas de caneta e são registradas as diferenças e semelhanças encontradas[11]. Assim, a diferença na composição ou na proporção dos elementos que constituem a mistura que forma a tinta da caneta é uma das formas pelas quais é possível se realizar a discriminação entre uma tinta e outra e, portanto, entre uma caneta e outra.

Vários métodos analíticos surgiram para diferenciar a composição das tintas de caneta. Alguns utilizam métodos de separação como cromatografia em camada delgada e eletroforese capilar para separar os componentes diferentes dos corantes e cromatografia líquida de alta performance e cromatografia gasosa acoplada à espectrometria de massas para identificar os componentes voláteis das tintas[11]. Entretanto, esses métodos não preservam a integridade da amostra e não são os mais indicados em análises forenses uma vez que essas lidam com amostras judiciais que podem ser objeto de reexame em caso de determinação judicial.

Calcerrada e García-Ruiz[8] apontaram que o uso de técnicas espectroscópicas, como espectrometria de massas, e espectroscopia na região infravermelha, Raman e ultravioleta e visível, tem sido crescente na análise de amostras forenses de documentos, inclusive, para análise de tintas de canetas. Nesse artigo de revisão, os autores demonstraram que a maior parte das pesquisas trata da discriminação de amostras de tintas pela identificação dos componentes principais de cada amostra e a subsequente comparação entre esses. Os autores destacam, ainda, que as técnicas de espectroscopia óptica têm se tornado de uso comum nos laboratórios por seu caráter não destrutivo, velocidade, simplicidade e fácil interpretação, mas que o fato das tintas poderem ser opticamente semelhantes pode levar a dificuldades na diferenciação das amostras.

A polícia brasileira comumente emprega técnicas baseadas em espectroscopia. Instrumentos mais simples, como o Comparador Espectral de Vídeo, são usados para o exame de documentos, sendo uma técnica não-destrutiva a qual, associada à aplicação de ferramentas quimiométricas, pode ser utilizada para a diferenciação de determinadas tintas de caneta[4, 5.]

2.2.1. Comparador Espectral de Vídeo

Um Comparador Espectral de Vídeo (do inglês, Video Spectral Comparator), VSC®, é o nome dado a uma variedade de instrumentos produzidos pela "Foster e Freeman" para auxiliar na detecção de falsificação ou adulteração de documentos[17]. O equipamento permite, por meio da combinação de imagem digital com radiação eletromagnética, a detecção de uma variedade de características que podem confirmar ou refutar a autenticidade do documento[17,18].

As Figuras 1e 2 trazem, respectivamente: a imagem da unidade principal do equipamento; um esquema representativo de funcionamento do aparelho, com destaque para a câmera de vídeo e para as fontes de iluminação.

Figura 1 - Unidade Principal do VSC® e Monitor do computador (adaptada do manual do equipamento)[17].

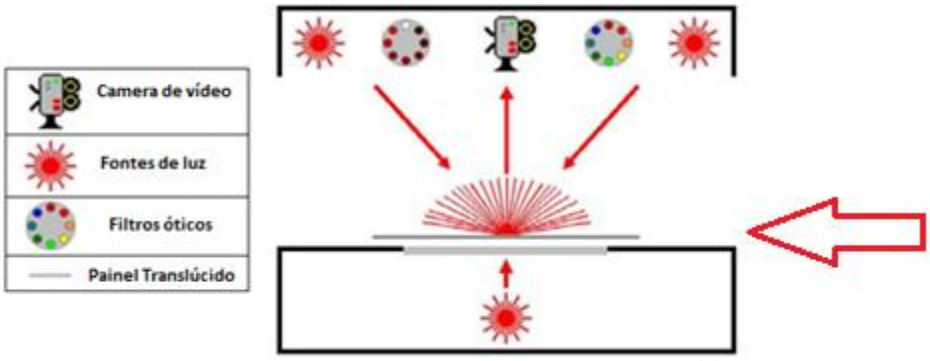


Figura 2 - Sistema de iluminação e de câmera do aparelho (adaptada do manual do equipamento) [17].

Conforme se observa na Figura 2, o documento a ser examinado é colocado dentro do equipamento, sobre uma placa de marquise: um painel de material translúcido no centro no qual o documento é posicionado (conforme apontado pela seta vermelha, na imagem), abaixo de onde são montadas as fontes de luz. Essas fontes de radiação podem fornecer iluminação nas regiões de ultravioleta, visível e infravermelho do espectro. No presente trabalho, utilizou-se a radiação do visível do espectro. A figura 2 traz o conjunto de lâmpadas do aparelho: a lâmpada utilizada na análise de imagens hiperespectrais é uma lâmpada halógena de filamento incandescente e está localizada na parte superior do aparelho.

Na função de espectrometria do instrumento - quando se incide radiação eletromagnética sobre o documento para analisá-lo – essa energia proveniente da radiação é dividida em várias partes: uma parte é refletida; uma parte, transmitida; uma parte, absorvida e uma parte é reemitida como fluorescência. Assim, a geração dos espectros correspondentes a cada um desses termos gera as modalidades disponíveis de espectros: reflectância, absorção, transmitância e fluorescência[18].

Um comparador espectral de vídeo é um instrumento relativamente simples de operar, não destrutivo e é eficaz na resolução de muitos casos investigados rotineiramente, desde que seja encontrado um comprimento de onda seletivo que diferencie as tintas [6]. Em análise de tintas de caneta, a associação do comparador espectral de vídeo à análise multivariada aumenta o potencial discriminante dessa técnica[4].

2.3. Métodos baseados em imagens associados à quimiometria

Na análise de documentos, quando as propriedades das cores das canetas que serão discriminadas forem muito semelhantes entre si, é possível que apenas a inspeção visual ou a utilização de imagens em escala RGB não seja suficiente para distinguir as canetas[7]. A fim de superar essas limitações, tem-se utilizado a combinação de informações químicas (espectrais) e espaciais, pelo sistema de imagens hiperespectrais.

Dados hiperespectrais são constituídos de uma imagem na qual se possui um espectro medido em cada pixel da imagem, gerando dados em três dimensões (pixels x pixels x comprimento de onda)[18]. Uma imagem hiperespectral, HSI (do inglês,

Hyperspectral Image) combina informações espaciais obtidas de imagens digitais e informações multiespectrais obtidas a partir de espectroscopia[6]. Os resultados são obtidos e apresentados sob a forma de um "Cubo hiperespectral", ou seja, dados tridimensionais, com duas dimensões espaciais e uma espectral[17,19], conforme se observa na Figura 3.

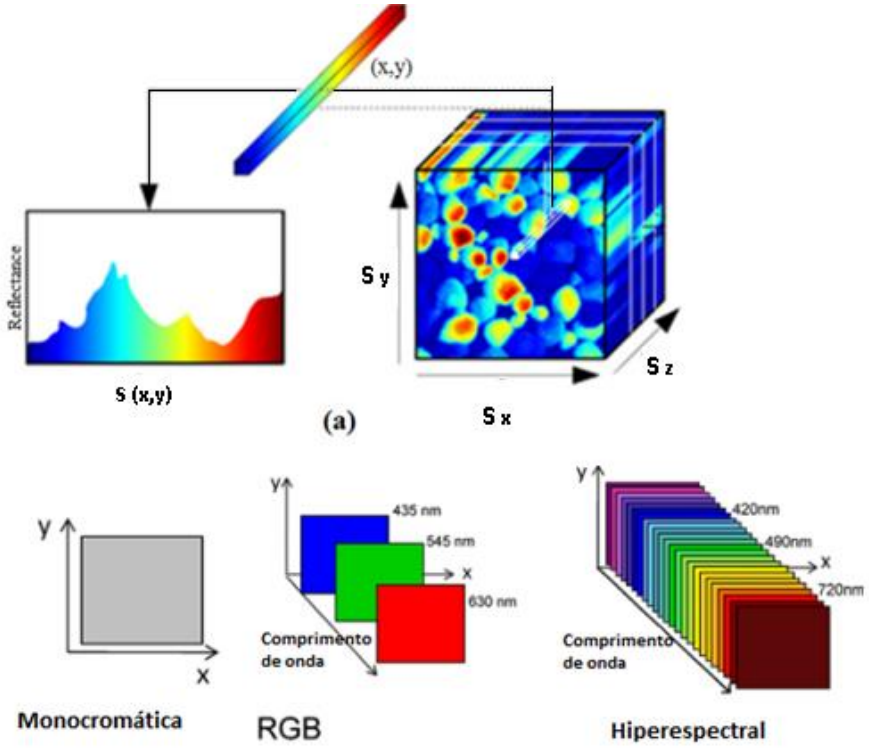


Figura 3 - Imagem hiperespectral: (**a**). Uma imagem hiperespectral representada no cubo 3D: um espectro para cada dimensão espacial (x,y). As figuras em sequência demonstram a formação, respectivamente, de uma imagem monocromática; de uma RGB e de uma hiperespectral[19,20].

A utilização de imagens hiperespectrais tem se mostrado crescente em diferentes especialidades. Edelman et al [13] demonstram várias aplicações dessa técnica em campos diversos como no controle de qualidade de produtos e de processos farmacêuticos e alimentares bem como o uso da técnica em diagnósticos médicos. Seu uso em química forense também se destaca com estudos em análise de impressões papilares, drogas, cabelos, manchas de sangue, preservativos, resíduos de explosivos e tintas, dentre outras aplicações[13]. A análise hiperespectral, pelo tamanho elevado dos dados que formam sua composição, está inerentemente vinculada à análise multivariada feita pela

quimiometria e seu grande uso não seria possível sem a implementação de algoritmos necessários para lidar com os dados gerados por uma única imagem[21.]

Em 2014, pela análise de tintas, Pimentel et al[14]demonstraram o uso de imagens hiperespectrais na região do infravermelho próximo, associado à quimiometria, com a utilização das técnicas de Análise de Componentes Principais, PCA (do inglês, *Principal Component Analysis*) e de Análise Multivariada de Curvas, MCR (do inglês, *Multivariate Curve Resolution*), em falsificação de documentos por adição de texto, por obliteração e por cruzamento de traços. Nesse estudo foi possível demonstrar o excelente potencial da técnica para detecção de falsificação de documentos, com a identificação da fraude em 82% nos casos de adição de texto e em 43% nos casos de obliteração, proporcionando um método mais objetivo e menos dependente de julgamentos pessoais do que os métodos tradicionais que requerem inspeção visual de uma pessoa qualificada e treinada para esse fim.

Posteriormente, em 2017, Pereira et al[6] analisaram o uso dessa técnica nas regiões de infravermelho próximo e médio, para casos de adição de texto associada à combinação de duas técnicas de quimiometria (PCA e *Projection Pursuit*) para discriminar tintas pretas e revelar possíveis adições e alterações de números, demonstrando o grande potencial de discriminação e exame objetivo da técnica, obtendo resultados, pela combinação das técnicas e das diferentes regiões de infravermelho, em que, em 90% das amostras de teste cego, foi possível se chegar à discriminação entre duas tintas ou entre papel e caneta.

A utilização da espectroscopia de imagem na região Raman para casos de obliteração, associada à Análise Multivariada de Curvas [7], também demonstrou bons resultados em que, em todos os cenários de obliteração estudados, a identidade das tintas e do texto obliterado foi satisfatoriamente recuperada.

Imagens hiperespectrais têm se mostrado muito útil na documentoscopia. O conjunto de dados gerado pelas análises de HSI é muito grande e, por isso, o uso da quimiometria se torna necessária para selecionar a informação química útil e melhorar a qualidade visual dos dados[6]. A quimiometria permite retirar informação de um conjunto de dados que pode apresentar uma dimensão que não permite mais a sua exploração por métodos estatísticos convencionais[17]. Nessa perspectiva, a Análise de Componentes Principais, PCA, é uma técnica exploratória amplamente utilizada para

análise de imagens, a qual não requer nenhum conhecimento prévio de classes ou tipologias nos dados, e permite a compactação de grandes conjuntos de elementos, bem como a seleção e a extração de características relevantes sem perda de informações valiosas[6]. A Análise Multivariada de Curvas, MCR, é uma técnica Quimiométrica destinada a resolver o problema de análise de misturas[22] e que também tem demonstrado bons resultados em dados de análise hiperespectral: por meio da análise por MCR é possível obter espectros puros dos constituintes que compõem a mistura analisada e recuperar os mapas de distribuição relacionados[23].

Associar o uso de HSI com um instrumento de aplicação consagrada na análise forense de documentos, como o VSC®, aliado à utilização de técnicas de quimiometria, pode representar uma ferramenta muito útil na análise objetiva de falsificações por adição de texto e por obliteração.

3. Fundamentação Teórica – Quimiometria

A quimiometria pode ser definida como uma disciplina da química que utiliza métodos matemáticos e estatísticos para planejar ou otimizar procedimentos experimentais e para extrair o máximo de informação química relevante, através da análise dos dados[24]. A quimiometria busca aplicar as ferramentas matemáticas e estatísticas adequadas que permitam retirar, do imenso conjunto de dados gerados por instrumentos computadorizados cada vez mais avançados, informações úteis para problemas de interesse da química[26].

Com o avanço da instrumentação, uma quantidade cada vez maior de dados vem sendo gerada e em diversos níveis de complexidade[25]. Passou-se a ter várias medidas sobre o sistema em análise e não apenas uma resposta para cada amostra, como era o observado na análise univariada, em que a influência de cada variável de interesse era estudada separadamente. Segundo Ferreira[26], para que uma análise de um conjunto de dados seja realizada de maneira adequada, são necessárias a organização e a preparação dos dados, uma análise exploratória desses e, de acordo com o objetivo final do estudo, a construção de modelos de classificação ou de regressão em estudos qualitativos e quantitativos, respectivamente.

3.1. Imagem hiperespectral e análise multivariada

A análise hiperespectral é inerentemente vinculada a dados multivariados[21]. Nesse caso obtém-se a mesma imagem em diferentes comprimentos de onda, de maneira que o conjunto de dados resultante é um bloco tridimensional de dados - chamado hipercubo -, com duas dimensões espaciais (x, y) e uma dimensão de espectral (z)[13], conforme se observa na Figura 3.

O objetivo final da análise hiperespectral é obter uma imagem que contenha informação seletiva e específica (quantitativa, grupos ou distribuição espacial) dos componentes presentes na superfície analisada[21]. Esse objetivo pode ser alcançado por meio da análise multivariada, utilizando as ferramentas quimiométricas adequadas.

O hipercubo fornece imagens para cada comprimento de onda, de forma que em cada pixel tem-se um espectro [13]. O cubo hiperespectral permite a visualização da distribuição dos componentes em uma amostra. Quando desdobrado em uma matriz de duas dimensões, esse conjunto de dados pode ser analisado com ferramentas estatísticas que permitam extrair, do grande conjunto de dados gerados, informações necessárias para resolução de problemas de interesse da química e, especificamente, de análise documentoscópica forense, como é o caso da presente pesquisa.

Uma proposta de análise quimiométrica de organização e de tratamento de dados para HSI é a descrita pela Figura 4: a partir da organização dos dados gerados pela resposta do instrumento utilizado nesta pesquisa – um comparador espectral de vídeo –, uma matriz bidimensional de dados é organizada; os dados são analisados com as ferramentas estatísticas adequadas e a informação de interesse químico é obtida.

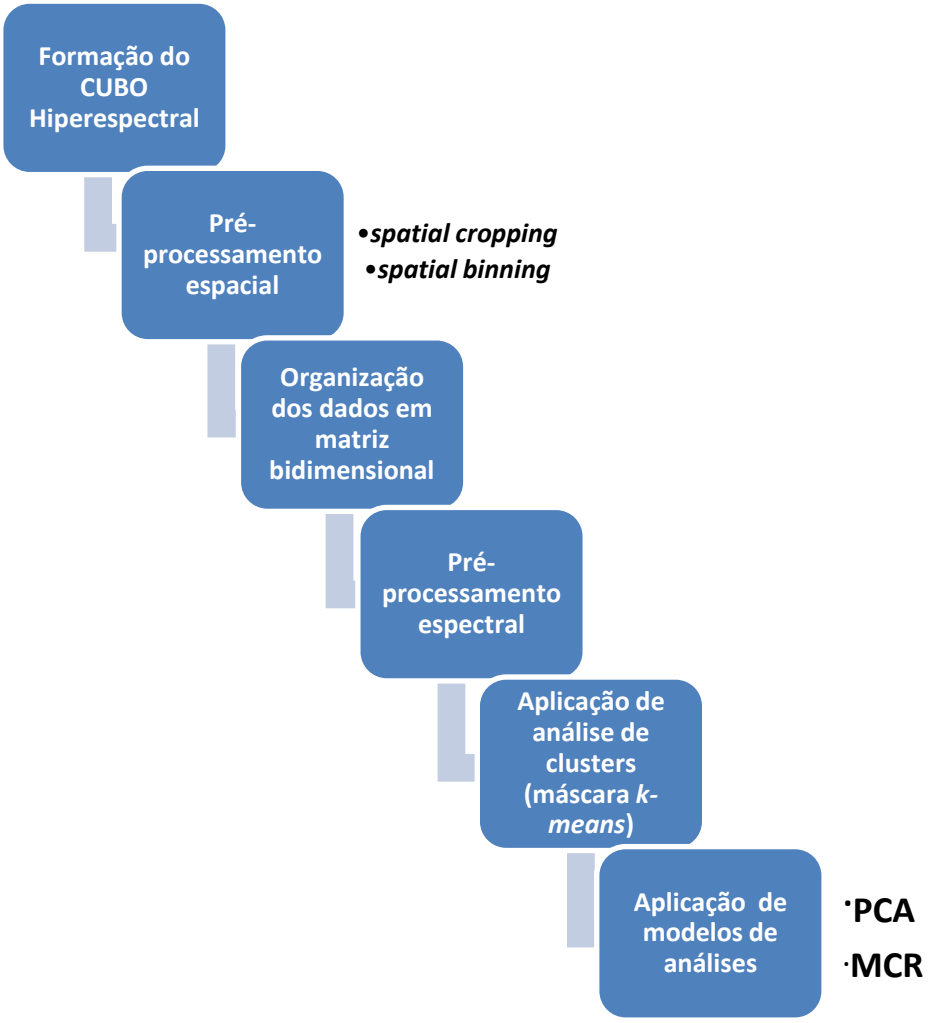


Figura 4: Proposta de tratamento quimiométrico para dados de HSI.

A proposta de análise quimiométrica executada neste trabalho é a abordada na Figura 4, com a utilização das técnicas consideradas adequadas para os dados experimentais utilizados nesta pesquisa, as quais serão explicadas a seguir:

3.1.1. Pré-processamento espacial

O pré-processamento espacial de imagens hiperespectrais é uma ferramenta opcional e tem como objetivo reduzir o tamanho das matrizes dos dados, eliminando as áreas da imagem que não sejam de interesse na análise multivariada[27]. No programa de análise de dados utilizado na presente pesquisa, há duas ferramentas utilizada para o pré-processamento espacial dos dados: são a "spatial cropped" a qual permite o corte da imagem, limitando a imagem apenas à região de interesse a ser analisada e a "spatial binning" a qual opera a redução de resolução da imagem, dividindo a escala original

em um número especificado pela janela: por exemplo, a aplicação da ferramenta "spatial binning", com uma janela de 5 pontos, reduz a escala original em 5 vezes, diminuindo a resolução da imagem.

3.1.2. Organização dos dados

Os dados hiperespectrais são organizados no formato de um cubo tridimensional: com duas dimensões espaciais (x, y) e uma dimensão de comprimento de onda (z)[13], conforme se observa na Figura 3. Esse hipercubo fornece imagens para cada comprimento de onda e um espectro pode ser obtido de cada pixel individual (x_j, y_k) [13]. O cubo hiperespectral é, portanto, uma matriz tridimensional com informações espaciais e espectrais. Quando desdobrado em uma matriz de duas dimensões, esse conjunto de dados pode ser tratado com as técnicas quimiométricas adequadas[14], conforme se observa na figura 5.

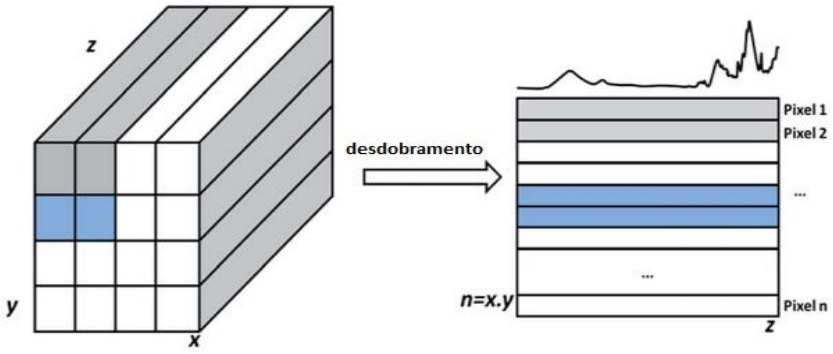


Figura 5– Desdobramento do cubo tridimensional (Cubo Hiperespectral) em matriz de dados bidimensional: **x** e **y** possuem a informação espacial e **z**, a informação química (ou espectral). Adaptado da referência 14.

3.1.3. Pré-processamento espectral

O objetivo do pré-processamento espectral é diminuir a influência de fenômenos indesejáveis que afetam a medição espectral, como o ruído instrumental[28]. Métodos de alisamento são utilizados com a finalidade de reduzir matematicamente o ruído, aumentando a relação sinal/ruído. Nesses métodos, seleciona-se uma janela, a qual contém um certo número de variáveis. Em seguida, ajusta-se um polinômio a esses pontos para determinar o valor no ponto central da janela. O tamanho da janela influencia diretamente o resultado do alisamento[29]. Diferentes algoritmos já foram propostos para a suavização ou cálculo de derivadas de espectros, mas o método de Savitzky-Golay, utilizando-se um polinômio de ordem baixa para determinar o valor

no ponto central da janela[29], é um dos mais conhecidos e utilizado[27]. A escolha do grau do polinômio afeta grandemente o resultado do alisamento, com polinômios de graus mais baixos removendo mais ruído [30].

3.1.4. Seleção de regiões específicas da imagem por agrupamentos por K-means

A análise de agrupamentos corresponde a métodos de classificação que realizam a análise da imagem segmentando os dados em uma composição química específica, com base na informação dos pixels[9]. Esses métodos de classificação se baseiam nas similaridades encontradas no conjunto de dados. Um exemplo de análise de agrupamentos é o realizado pelo algoritmo k-means em que se atribui cada pixel da imagem a um agrupamento (*cluster*) de ordem "k", cujo centro é o mais próximo, minimizando a soma dos quadrados das distâncias de cada pixel ao seu centro correspondente. A principal vantagem deste algoritmo é sua simplicidade de maneira a permitir sua aplicação em conjuntos maiores de dados [9]. Esse algoritmo é utilizado para imagens hiperespectrais com o intuito de selecionar de forma mais eficiente uma região de interesse aplicando uma "máscara". Um exemplo da aplicação desse método na análise de documentos é a seleção de clusters que contenham apenas tinta de forma a excluir da análise os pixels que apenas contenham papel.

3.1.5. Análise de Componentes Principais – PCA

Uma das áreas mais utilizadas na quimiometria é a de reconhecimento de padrões, por meio da qual podem ser encontrados agrupamentos e tendências em um conjunto de amostras [26]. O reconhecimento de padrões busca identificar as semelhanças e as diferenças nas amostras a fim de agrupá-las e classificá-las. Na quimiometria, os métodos de reconhecimento de padrões se dividem em "métodos supervisionados" e em "métodos não supervisionados" [26]. Nos métodos supervisionados é necessário que exista alguma informação inicial sobre a identidade das amostras para a formação das classes e o objetivo é desenvolver um modelo baseado nas informações contidas nas amostras. Já nos métodos não supervisionados, a separação de classes acontece sem a necessidade de informações iniciais sobre a natureza das amostras e o objetivo é identificar agrupamentos naturais entre as amostras [29, 31].

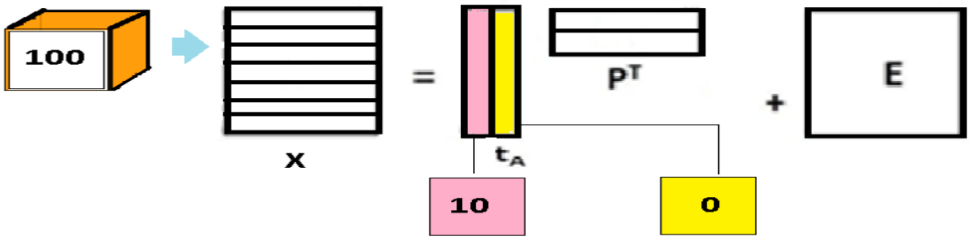
Nos métodos não-supervisionados, também conhecidos como "métodos de análise exploratória de dados", as amostras são agrupadas naturalmente com base nas

informações contidas nos dados experimentais obtidos [26]. Um dos objetivos da análise exploratória é a redução do número de variáveis dos dados para um conjunto que possa mais facilmente visualizado[14], permitindo a interpretação de conjuntos de dados complexos por meio de gráficos bi ou tridimensionais[29].

Um dos métodos matemáticos de análise exploratória que pode ser aplicado para a verificação dessas similaridades é Análise de Componentes Principais (PCA, do inglês *Principal Component Analysis*). A PCA é utilizada para visualizar a estrutura dos dados, encontrar similaridades entre amostras e reduzir a dimensionalidade do conjunto de dados. A análise exploratória através da PCA, além de ser um método de reconhecimento de padrões, também é a base para diversos métodos de classificação[29].

A ferramenta PCA permite a redução da dimensionalidade do conjunto de dados através da representação em um novo sistema de eixos, chamados de componentes principais (PC), permitindo a visualização da natureza multivariada em poucas dimensões [29]. A dimensionalidade do espaço original é reduzida sem que as relações entre as amostras sejam afetadas [26]. No espaço original, as amostras são pontos localizados em um espaço J-dimensional, "J" igual ao número de variáveis. Para "J" igual a quatro, por exemplo, teríamos um espaço dimensional com quatro dimensões – menos usual do que aqueles de representações cartesianas de até três eixos. Com a redução de dimensionalidade operada pela PCA, as amostras passam a ser pontos localizados em espaços de dimensões reduzidas definidos pelas PCs, por exemplo, bi- ou tridimensionais [29].

Em termos matemáticos, na Análise de PCA, a matriz original **X** é decomposta em um produto de duas matrizes, denominadas escores (**T**) e pesos (**P**), mais uma matriz de erros (**E**). A análise de PCA pode ser aplicada a um diverso conjunto de dados, inclusive a dados de origem hiperespectral, conforme ilustra a figura 6, em que o cubo tridimensional de dados hiperespectrais é desdobrado em uma matriz bidimensional a qual se aplica a ferramenta matemática PCA:



$$\mathbf{X} = \mathbf{TP^t} + \mathbf{E} \ (1)$$

Figura 6: Representação esquemática de Análise de PCA com imagem hiperespectral: a matriz original X é decomposta em um produto de duas matrizes, denominadas escores (T) e pesos (P), mais uma matriz de erros (E).

Os escores representam as coordenadas das amostras no sistema de eixos formados pelas componentes principais. Cada componente principal é constituída pela combinação linear das variáveis originais (quando se realiza combinações lineares das variáveis originais, estamos agrupando aquelas que fornecem informações semelhantes, definindo um novo conjunto de variáveis ortogonais com propriedades desejáveis e específicas). As novas variáveis são as PCs [26] e os coeficientes da combinação são denominados pesos [29]. Os pesos são os cossenos dos ângulos entre as variáveis originais e as componentes principais (PC), e representam, portanto, o quanto cada variável original contribui para uma determinada PC. Os escores representam as relações de similaridade entre as amostras. A avaliação dos pesos permite compreender quais variáveis mais contribuem para os agrupamentos observados no gráfico dos escores. Na representação da figura 6, o conjunto de dados hiperespectrais possui informações relativas ao número “100”, supostamente originado de duas tintas diferentes, uma para o número dez (10) e outra para o segundo zero (0). Esse conjunto original de variáveis é reduzido pela combinação linear e forma-se um novo conjunto de variáveis nas quais se espera que os dados que descrevem o número 100 possam ser descritos em termos de novas variáveis que separem os dados em dois números, 10 e 0, capazes de demonstrar uma informação que já estava presente nos dados originais, mas que, pela aplicação da técnica de PCA, torna-se mais facilmente visualizada.

Na formação das componentes principais, a primeira componente principal (PC1) é traçada no sentido da maior variação no conjunto de dados; a segunda (PC2) é traçada ortogonalmente à primeira, com o intuito de descrever a maior porcentagem da variação não explicada pela PC1 e assim por diante. Uma informação que está contida

em uma das componentes principais não está presente em outra (isso significa que elas são "não" correlacionadas e são ortogonais entre si). Pela própria maneira como essas novas variáveis são definidas, uma vez que as redundâncias são removidas, é possível descrever quase toda a informação contida nos dados originais usando apenas algumas e poucas componentes principais [26]. Por meio da análise conjunta do gráfico de escores e de pesos, é possível perceber quais variáveis são responsáveis pelas diferenças observadas entre as amostras [29].

Existem muitos algoritmos disponíveis para o cálculo das matrizes de pesos e de escores. A Decomposição por Valores Singulares, SVD (do inglês, *Singular Value Decomposition)*, é o método cuja técnica numérica é considerada mais acurada e estável para o cálculo das componentes principais[26] e foi o utilizado para o cálculo das componentes principais nos modelos aplicados no presente trabalho.

O número de componentes principais pode ser determinado em função da porcentagem de explicação da variação presente no conjunto de dados originais [29,32].

3.1.5.1. Centragem de dados na média

Esse é o tipo de processamento adequado para eliminar eventuais ruídos provenientes de dados experimentais e indicado quando as variáveis possuem a mesma natureza e são obtidas da mesma fonte [16]. Para realizar a operação matemática, primeiro calcula-se o valor médio de cada coluna da matriz de dados e, então, subtrai-se esse valor de cada um dos valores da respectiva coluna, conforme mostra a equação 2 em que $\bar{x}_j$ é a média aritmética dos valores de uma determinada variável *j*, na presença de *I* amostras; x_{ij} é o valor da variável *j* na amostra i e $x_{ij(CM}$ é o dado centrado na média para a variável *j* na amostra *i*. A operação realizada por essa ferramenta apenas desloca a origem dos eixos [16], fazendo uma translação de eixos para o valor médio de cada um desses de forma que a estrutura dos dados é completamente preservada [26].

$$x_{i,j}(\mathbf{cm}) = x_{i,j} - \bar{x}_j \quad (2)$$

em que $\bar{x}_j = \frac{1}{I}\sum x_{i,j}$ é $a\ média\ da\ j-ésima\ coluna\ dos\ dados.$

3.1.6. Análise Multivariada de Curvas – MCR

O modelo de Resolução Multivariada de Curvas (MCR, do inglês Multivariate Curve Resolution) têm o intuito de resolver misturas de sinais[23] e têm a sua expressão matemática semelhante à Análise de Componentes Principais. A Resolução Multivariada de Curvas é uma técnica que utiliza a combinação linear para formação de novas variáveis, análogas aos pesos de PCA, mas que se tem uma maior capacidade de interpretação no sentido químico, pois devem se aproximar do espectro puro de cada componente. A decomposição de uma matriz de dados hiperespectrais, por meio da técnica de MCR, promove a descrição da mistura inicial em termos de novas variáveis relacionadas à concentração de seus componentes puros. Em uma mistura de tintas, como a analisada na presente pesquisa, a decomposição da matriz de dados será feita em escores, que possuem um significado físico relacionado à intensidade relativa dos componentes da tinta e os pesos, que devem ser estimativas dos espectros de cada componente da tinta.

Para inicializar o algoritmo de análise por MCR, é necessária uma estimativa inicial dos espectros ou das concentrações dos componentes puros presentes na matriz de dados. Essas estimativas são obtidas por métodos baseados na aproximação da variável pura. Esses métodos selecionam as colunas com as variáveis mais puras de acordo com o número de fatores que se acredita existirem na amostra[23]. O método utilizado na presente pesquisa, o PURITY, realiza as estimativas iniciais dessa forma.

Para que o método tenha resultados mais condizentes com as informações químicas[23], utilizam-se algumas restrições. Uma delas é a restrição de não-negatividade, utilizada na presente pesquisa. A não-negatividade impõe que os sinais dos espectros e os perfis de concentração sejam sempre positivos.

4. Materiais e métodos

A parte experimental desta pesquisa foi desenvolvida no Laboratório de documentoscopia do Serviço de Perícias Documentoscópicas do Instituto Nacional de Criminalística do Departamento de Polícia Federal, em Brasília.

4.1. Produção das amostras

Para o desenvolvimento de toda a pesquisa, foram adquiridas 17 canetas, esferográficas de cor azul, de marcas e modelos distintos, no comércio local do Distrito Federal. As amostras são lançamentos de tintas de caneta em papel, produzidas pela pesquisadora, simulando situações reais para os casos de fraude estudados: por adição de texto e por obliteração.

Todas as amostras foram produzidas em folha de papel branco, do tipo A4, com gramatura de 75 g/m^2, da marca Report Suzano. Após a produção, todos os padrões de amostras foram guardados longe de possíveis agentes de degradação por luminosidade, acondicionados em sacos plásticos e colocados em uma caixa de papel, sem iluminação e à temperatura ambiente.

4.1.1. Adição de texto

Para a seleção das canetas e formação das amostras de adição de texto, foram utilizadas combinações que não fossem distinguíveis a olho nu, a fim de gerar amostras mais verossímeis: a figura 7 demonstra um exemplo de combinação. A Tabela 1 traz o conjunto total de canetas que compreende 17 unidades, do tipo esferográfica, de cor azul, de marcas e modelos distintos.

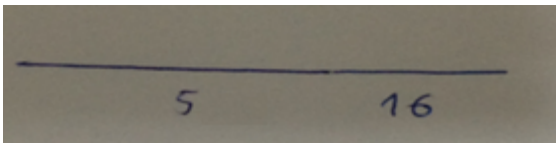


Figura 7– Imagem obtida por smartphone da combinação de traços com canetas diferentes.

Tabela 1. Relação das marcas e modelos de canetas utilizadas para as amostras de adição de texto.

Canetas utilizadas - Adição de Texto			
Tipo de Caneta	**Marca**	**Modelo**	**Numeração da caneta**
Esferográfica (E)	***Paper Mate***	Kilometrica 300RT 1.0M	**1**
	Pilot	BP-S	**2**
	Pilot	BP-1 Inox 0.7	**3**
	Maped	Ice	**4**
	Tilibra	Definit 0.6	**5**
	Schneider	Slider XB	**6**
	Pilot	Super Grip 0.7	**7**
	Pilot	Super Grip 1.0	**8**
	Cis	BPM-01	**9**
	Cis	Super Bold 1.2 mm	**10**
	Uni	Lakubo 0.7	**11**
	BIC	Cristal	**12**
	Faber Castell	Trilux 032 Medium	**13**
	Pilot	Super Grip 1.6	**14**
	Pilot	BP-1 Inox 1.0	**15**
	Uni	Laknock II fine	**16**
	Compactor	O7	**17**

As amostras foram produzidas simulando a adulteração do texto pelo acréscimo de um número: transformou-se um número “10 ,0” em um número “10 0,0” pela adição de um texto – o segundo algarismo “0” escrito antes da vírgula. O texto original – o número “10 ,0” – foi escrito com uma caneta (caneta identificada como caneta 1) e o trecho adicionado - o segundo algarismo “0” - foi escrito com outra caneta (caneta identificada como caneta 2). Assim, uma amostra de adição de texto é sempre produzida em um papel pela combinação de duas canetas diferentes. A Tabela 2 traz um exemplo de produção das amostras, para a amostra número 22.

Tabela 2. Exemplo de produção amostras – Adição de texto.

Caneta 1	Caneta 2	Descrição do que foi escrito com cada caneta
1	11	10 ,0 (1)
		0 (2)

Ao todo, foram produzidas 50 amostras de adição de texto, conforme descrita na Tabela 3.

Tabela 3. Total de amostras de adição de texto.

Número da amostra	Combinação de canetas utilizadas	Número da amostra	Combinação de canetas utilizadas
1	16 e 1	26	1 e 12
2	16 e 13	27	1 e 2
3	16 e 6	28	9 e 7
4	16 e 15	29	9 e 10
5	16 e 7	30	9 e 8
6	16 e 4	31	9 e 14
7	16 e 10	32	9 e 17
8	16 e 11	33	9 e 2
9	16 e 8	34	13 e 9
10	16 e 14	35	13 e 15
11	16 e 17	36	13 e 7
12	16 e 3	37	13 e 10
13	16 e 12	38	13 e 8
14	16 e 5	39	13 e 14
15	16 e 2	40	13 e 17
16	1 e 13	41	13 e 3
17	1 e 6	42	13 e 12
18	1 e 15	43	13 e 2
19	1 e 7	44	10 e 6
20	1 e 4	45	10 e 15
21	1 e 10	46	10 e 7
22	1 e 11	47	10 e 11
23	1 e 8	48	6 e 14
24	1 e 14	49	6 e 17
25	1 e 3	50	6 e 3

4.1.2. Obliteração

De modo semelhante às amostras de adição de texto, para a produção das amostras de obliteração, foram utilizadas canetas cujas combinações não fossem distinguíveis a olho nu. A figura 8 demonstra um exemplo de combinação.

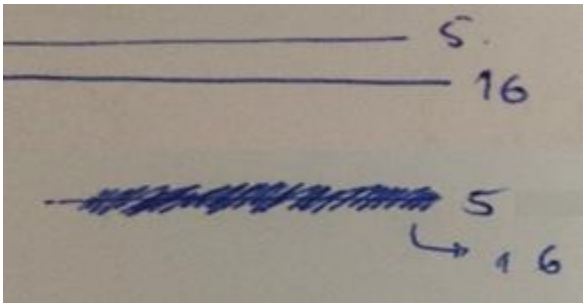


Figura 8 –Imagem obtida por smartphone da combinação de traços com canetas diferentes – Obliteração (a caneta 5 foi escrita primeiro e obliterada pela caneta 16).

A Tabela 4 traz a relação total de canetas utilizadas: 6 unidades, do tipo esferográfica, de cor azul, de marcas e modelos distintos.

Tabela 4. Relação das marcas e modelos de canetas utilizadas para as amostras de obliteração.

Canetas utilizadas - Obliteração			
Tipo de Caneta	**Marca**	**Modelo**	**Numeração da caneta**
Esferográfica (E)	***Paper Mate***	Kilometrica 300RT 1.0M	**1**
	Tilibra	Definit 0.6	**5**
	Uni	Lakubo 0.7	**11**
	BIC	Cristal	**12**
	Uni	Laknock II fine	**16**
	Compactor	O7	**17**

As amostras foram produzidas da seguinte forma: escreveu- se uma palavra com uma caneta e após um intervalo de tempo – mínimo de 48 horas – cobriu-se a palavra com outra caneta. Assim, cada amostra de obliteração é formada pela combinação das tintas de duas canetas: caneta identificada como "caneta 1" – utilizada para escrever o texto; caneta identificada como "caneta 2" – utilizada para cobrir o texto. O texto escrito é "FALSO" e é o mesmo texto em todas as amostras.

Assim, uma amostra de obliteração é sempre produzida em um papel pela combinação de duas canetas diferentes e a palavra obliterada é sempre “FALSO”. A Tabela 5 traz um exemplo de produção de amostras, para a amostra número 8.

Tabela 5. Exemplo de produção de amostras – obliteração.

Caneta 1	Caneta 2	Descrição do que foi escrito com a caneta 1
5	16	“FALSO” (1)

Ao todo, foram produzidas 15 amostras de obliteração, conforme descrita na Tabela 6.

Tabela 6. Total de amostras de obliteração

Número da amostra	Combinação de canetas utilizadas
1	1 e 5
2	1 e 11
3	1 e 12
4	1 e 16
5	1 e 17
6	5 e 11
7	5 e 12
8	5 e 16
9	5 e 17
10	11 e 12
11	11 e 16
12	11 e 17
13	12 e 16
14	12 e 17
15	16 e 17

4.2. Aquisição dos espectros no VSC 6000®

Foram adquiridos espectros de reflectância difusa utilizando o equipamento VSC 6000® (Foster e Freeman), na região de 400 a 1000 nm, com intervalos de medida de 4 nm, com o emprego da função hiperespectral do aparelho. O instrumento foi calibrado com o mesmo intervalo espectral (400 a 1000 nm) com um papel A4 branco

de mesma gramatura do utilizado nas amostras. A fonte de luz empregada foi uma lâmpada de halogênio, de 100 W de potência.

Realizou-se o branco da análise em uma região da amostra livre de tinta, isto é, o próprio papel, antes das medidas. A aquisição dos espectros foi na área contendo a tinta de caneta, com a imagem ampliada com o ajuste do zoom de maneira em que todo o texto que é objeto da adulteração (o algarismo adicionado, no caso de adição de texto; todo o texto oculto, no caso de obliteração) permanecesse dentro da imagem resultante da ampliação, com o uso do foco automático do aparelho.

Para adição de texto e para a obliteração, a ampliação foi de 7,56x. Foi adquirida uma imagem hiperespectral para cada amostra a qual foi convertida do formato de extensão "HSI" para o formato "CSV" por um programa fornecido pela Foster e Freeman. A resolução da imagem foi de 1292x978 pixels e 151 comprimentos de onda, para todas as imagens obtidas. Cada conjunto de dados gerado por amostra foi de, aproximadamente, 1,1 GB.

Após a medida, todos os padrões de amostras foram acondicionados em sacos plásticos e colocados em uma caixa de papel, sem iluminação e à temperatura ambiente.

4.3. Análise dos dados

Após a conversão para o formato adequado (.CSV), os dados foram exportados para o programa Matlab® (Versão R2018a), e analisados com o uso do pacote *Image Processing Toolbox*™. Utilizou-se também o HYPER-tools, versão 1.1, uma ferramenta de interface gráfica desenvolvida especificamente para análise de HSI[21]. A figura 9 descreve, para os casos de adição de texto e para obliteração, a aplicação das técnicas de análise com uma abordagem univariada e multivariada e a figura 10 traz a descrição das ferramentas para tratamento de dados na abordagem univariada e multivariada.

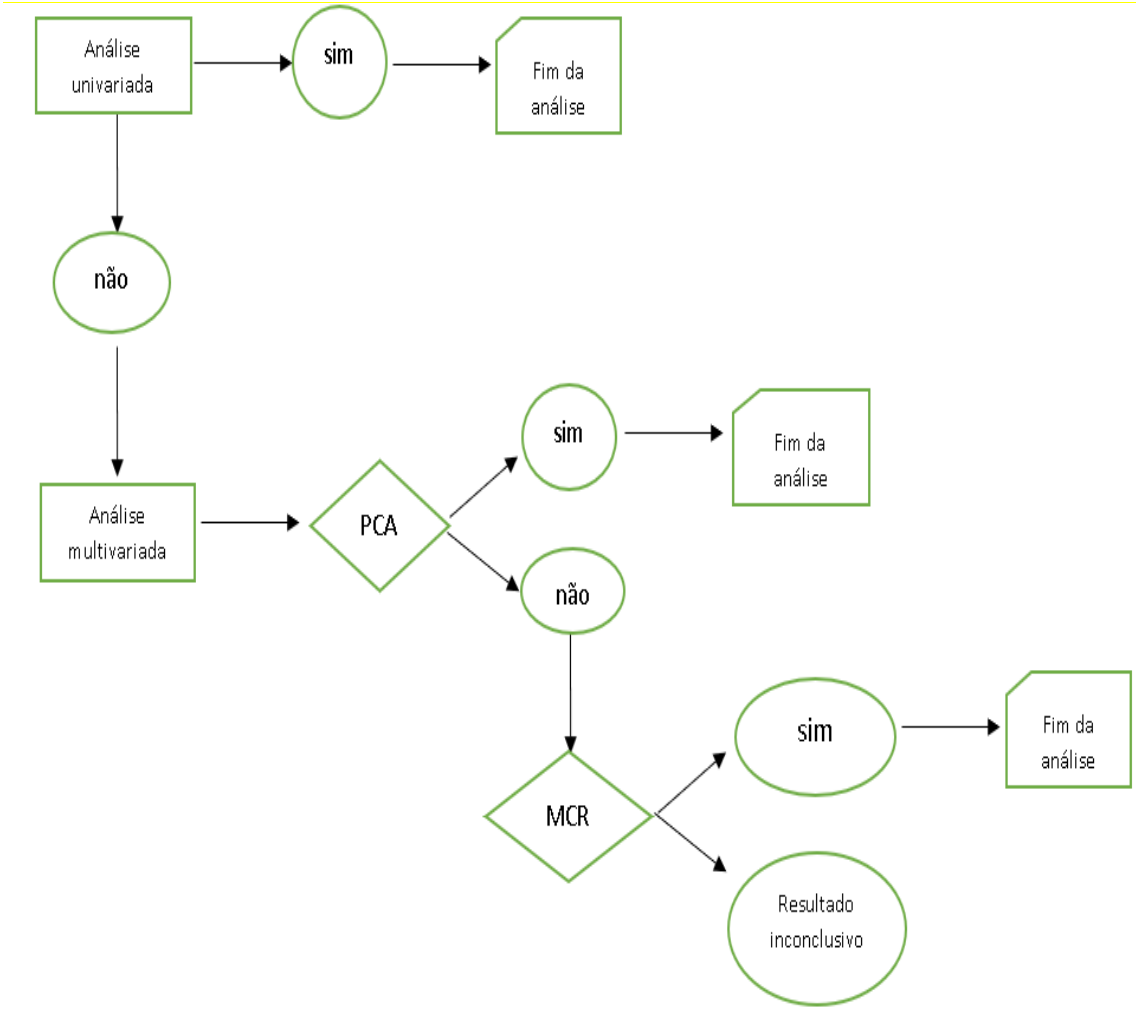


Figura 9 - Fluxograma 1 – Aplicação das técnicas de análise com uma abordagem univariada e multivariada.

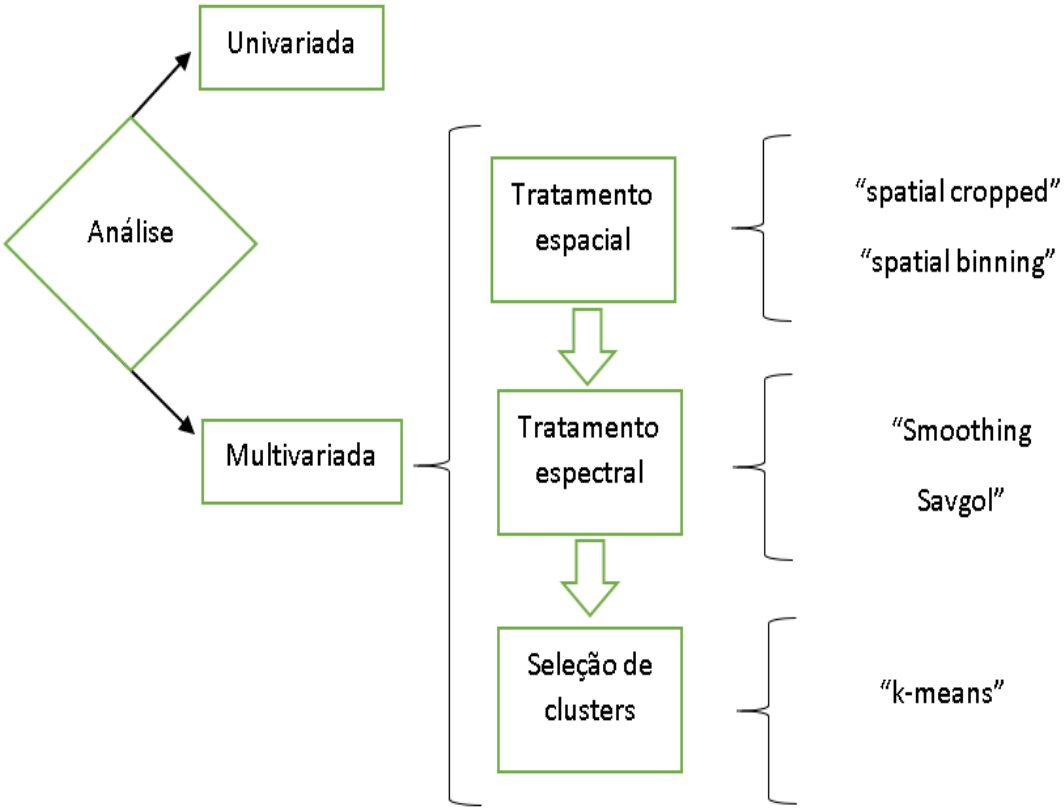


Figura 10 - Fluxograma 2 – Descrição das ferramentas para tratamento de dados na abordagem univariada e multivariada.

Nas análises de obliteração, utilizou-se um modelo semi-quantitativo em que, em função da quantidade de letras identificas na análise, prosseguia-se para a próxima técnica. O objetivo da técnica é revelar as letras da palavra "FALSO". Para fins de análise forense, toda informação obtida a partir da identificação de qualquer letra da palavra pode ser considerada útil para identificar a fraude de um documento. Assim, a análise tomou como critério a identificação da quantidade de letras, sendo a identificação correta de uma letra correspondente à taxa de 20%, a de duas letras, à taxa de 40%, à de três letras, à taxa de 60%, à de quatro letras, à taxa de 80% e a de cinco letras, à taxa de 100%.

Caso fosse possível revelar uma quantidade de letras, com uma variação de 20 a 100 % de taxa de identificação correta, apenas com o uso da análise univariada, não se prosseguia para a análise multivariada. Quando a análise univariada não se demonstrava suficiente, prosseguia-se para a análise multivariada. A análise era feita inicialmente por PCA e, de maneira semelhante, caso essa técnica não se demonstrasse

suficiente para a revelação dos caracteres, prosseguia-se, por fim, à análise por MCR. Se, com a utilização de todas as técnicas, não fosse obtida a revelação de nenhuma letra, a análise seria considerada inconclusiva.

4.4. Teste cego

Para os casos de obliteração, realizou-se um teste cego, pela aluna da presente pesquisa, identificada como pesquisadora, da seguinte forma: a pesquisadora informou que se tratava de uma pesquisa com análise de documentos e que, para a realização dessa, escreveu-se um texto, com sentido lógico (afirmou-se apenas que a imagem poderia conter algum caractere numérico ou alfanumérico), com uma caneta e, após, escondeu-se esse texto com a sobrecarga de outra caneta, passando-se a tinta de uma caneta sobre outra. A pesquisadora relatou aos entrevistados, pessoas identificadas como "leigos" que não tiveram treinamento prévio para a interpretação das imagens e que não tinham conhecimento do texto que estava escrito, que o objetivo do método era revelar o texto que havia sido escondido e que, se o método fosse eficaz, seria possível perceber uma sequência lógica que revelasse o texto. A pesquisadora informou que não existiam respostas certas ou erradas e que, seriam apresentadas imagens e que, dentre essas, o entrevistado deveria escolher aquela em que ele era capaz de identificar algum padrão de cores suficiente para identificar partes de um texto que fizesse sentido lógico. O conteúdo do texto não foi informado aos entrevistados.

Após as instruções, a pesquisadora apresentou as imagens obtidas pela análise univariada e, em função da quantidade de letras identificadas corretamente, a pesquisadora apresentou as imagens obtidas da análise por PCA, inicialmente, e por MCR, posteriormente, caso as técnicas anteriores não fossem eficazes para obter alguma informação relativa ao texto.

5. Resultados e discussões

Os resultados foram organizados em função do tipo de fraude analisada: adição de texto ou obliteração. De acordo com a metodologia adotada na presente pesquisa, sempre se iniciou a análise com uma abordagem univariada, buscando-se a seleção de um comprimento de onda em que seja possível a discriminação das tintas analisadas, e, quando essa não foi eficaz, utilizou-se a análise multivariada: inicialmente com PCA e, nos casos em que essa técnica não permitiu um resultado conclusivo, prosseguiu-se para o uso da técnica de MCR.

A fim de representar as situações observadas nas 50 amostras de adição de texto e 15 de obliteração, foram selecionadas algumas amostras representativas das situações e dos resultados encontrados. Organizou-se a apresentação desses resultados em função do tipo de método empregado para a discriminação das tintas, apresentando ao menos um caso em que foi aplicada a análise univariada; a análise multivariada – PCA ou MCR). Foi apresentada uma descrição detalhada das análises feitas em uma amostra, demonstrando os tratamentos dos dados e as ferramentas aplicadas no software HYPERTools no ambiente Matlab; uma descrição mais objetiva de outra amostra analisada e uma compilação final com todas as amostras analisadas e os resultados obtidos.

Demonstrou-se também a análise feita para os casos em que o resultado foi considerado inconclusivo, isto é, nenhuma das abordagens teve êxito em resolver o caso de adição de texto ou obliteração.

5.1. Adição de texto

As amostras de adição de texto são sempre produzidas simulando a falsificação de um número “100,0” formado pela adição de um texto, um “0”, a um trecho original “10”. Assim, a caneta identificada como a tinta 1 se refere à caneta do primeiro zero à esquerda e a caneta identificada como tinta 2 se refere ao segundo zero e todos esses números estão antes da vírgula.

- **Análises Univariada**

A amostra tomada como exemplo foi a amostra de número 6 na qual o primeiro "0", identificado como tinta 1, foi escrito com a caneta 16 e o zero adicionado, identificado como tinta 2, foi escrito com a caneta 4.

Após a aquisição dos dados do VSC®, esses são importados para o ambiente Matlab, organizados no formato do cubo hiperespectral e o cubo é desdobrado em uma matriz de dados. Nessa etapa, a matriz é formada de forma a selecionar apenas os dados relativos aos números "0" escritos com as canetas distintas. Cabe destacar que essa abordagem univariada também pode ser realizada no software do próprio equipamento, sendo que os resultados obtidos de forma univariada nesse software e no HYPERTools em ambiente Matlab são equivalentes.

Pela formação do cubo hiperespectral, os dados trazem informações relativas à imagem (medidas em pixel) e à composição química (medida em pseudo-absorbância para as variáveis de comprimento de onda). Para cada imagem, é possível realizar a correspondência a um comprimento de onda. Por meio da combinação dos dados relativos às imagens com a variação de um comprimento de onda por vez, tem-se a análise univariada na qual se seleciona um comprimento de onda em que seja possível discriminar as tintas, combinando-se a informação de imagem com a do espectro. Assim, é possível diferenciar as tintas na imagem pela seleção de um comprimento de onda. O software permite fazer uma varredura ao longo dos 151 comprimentos de onda, fornecendo a imagem, composta por um conjunto de pixels, correspondente a cada um dos comprimentos de onda. Quando é possível discriminar as tintas na imagem pela seleção de um comprimento de onda tem-se pela análise univariada uma imagem que evidencia que as tintas são diferentes através do desaparecimento ou nítida diferença de intensidades de um dos zeros da imagem em relação ao outro no comprimento de onda escolhido. A figura 11 demonstra um exemplo de análise em que foi possível discriminar os dois zeros com a seleção do comprimento de onda 704 nm.

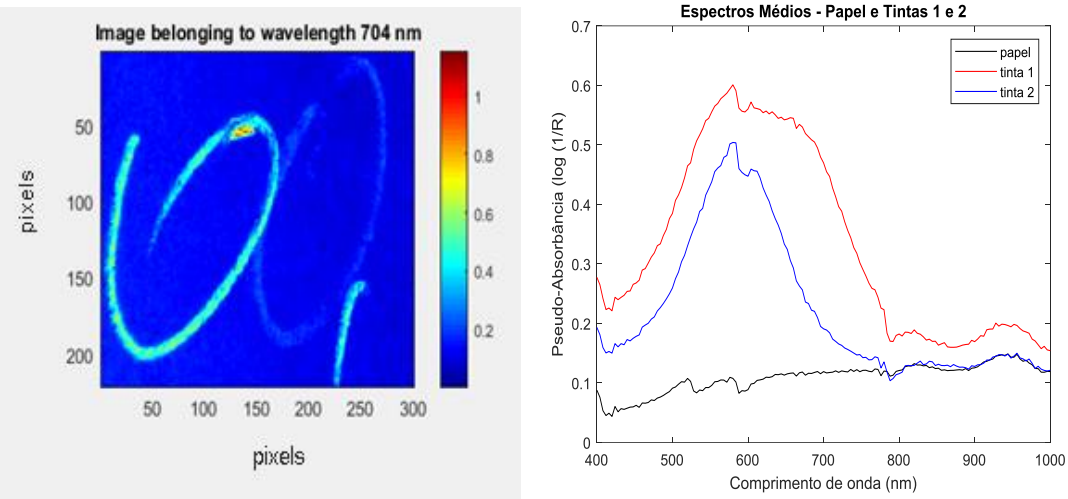


Figura 11 – À esquerda, imagem correspondente ao comprimento de onda de 704 nm e à direita, gráfico dos espectros médios de pixels do papel, da tinta 1 e da tinta 2 – ambos da amostra 6 de adição de texto.

O resultado demonstra que foi possível encontrar um comprimento de onda seletivo para essa amostra (704 nm), que foi capaz de diferenciar o zero escrita pela tinta 1 do zero escrito pela tinta 2. Ou seja, esse comprimento de onda é seletivo para uma das tintas de maneira a diferenciá-las na imagem e apenas com a variação de uma variável é possível obter uma resposta que diferencie a tinta das canetas utilizadas. A resposta é a figura da imagem hiperespectral no comprimento de onda selecionado. A escala de cores na imagem da figura 11 corresponde à intensidade do espectro obtido em pseudo-absorbância e é crescente da cor azul para a vermelha. O gráfico da figura 11 demonstra que foi possível verificar espectros distintos para a tinta 1, para a tinta 2 e para o papel.

Um segundo exemplo em que se verifica a diferenciação dos dois zeros escritos com tintas diferentes é o da amostra 22 de adição de texto, na figura 12: o primeiro zero, relativo à tinta 1, foi escrito com caneta 1 e o zero adicionado - relativo à tinta 2 – foi escrito com a caneta 11. Percebe-se que a imagem selecionada para o comprimento de onda 724 nm diferencia as duas tintas.

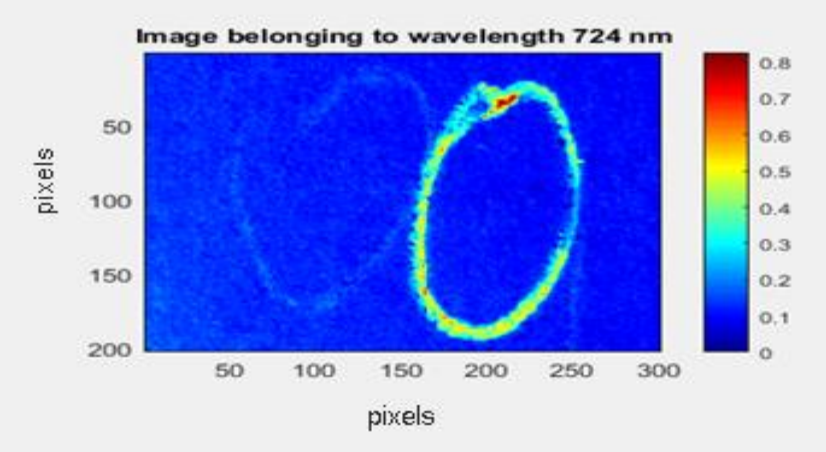


Figura 12 – Imagem, em pixels, correspondente ao comprimento de onda 724 nm - amostra 22 de adição de texto.

Ao todo, foi possível, por meio da análise univariada, diferenciar o trecho adicionado do trecho original, em 25 das 50 amostras produzidas, representando 50%, do total de amostras produzidas.

- **<u>Análises Multivariada</u>**

Análise de Componentes Principais – PCA

A amostra tomada como exemplo foi a amostra de número 15 na qual o primeiro "0", identificado como tinta 1, foi escrito com a caneta 16 e o zero adicionado, identificado como tinta 2, foi escrito com a caneta 2.

De acordo com o método de análise, com os dados organizados no formato do cubo hiperespectral e tratados por meio do programa HYPERTools. Inicialmente procede-se à análise univariada, buscando-se um comprimento de onda que seja seletivo para uma das tintas, e uma vez que a técnica não se demonstre eficaz, utiliza-se a análise multivariada, iniciando-se por PCA. A figura 13 demonstra a imagem obtida para a análise univariada da amostra e percebe-se que não há distinção tão visível dos pixels capaz de discriminar um zero do outro: a imagem não demonstra uma diferença de tom de cor tão nítida entre os dois zeros. O equipamento faz uma varredura de análise ao longo da faixa espectral de 400 a 1000 nm, trazendo várias imagens, em pixels, mas nenhuma delas diferenciou os zeros das duas canetas. A imagem correspondente à figura 13 demonstra um exemplo representativo de uma das imagens obtidas na varredura.

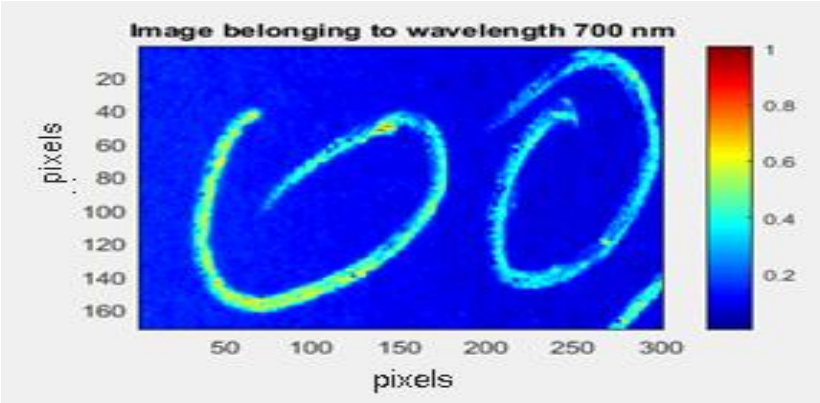


Figura 13 – Imagem, em pixels, correspondente ao comprimento de onda 700 nm- Amostra 15 de adição de texto.

No prosseguimento da análise, após a verificação de que não foi possível a discriminação por análise univariada, os dados são analisados a fim de se proceder à PCA: utiliza-se o pré-tratamento espacial - para redução do tamanho dos dados– com a aplicação de uma ferramenta para selecionar a região de interesse da imagem (*spatial cropping)* – eliminando-se as áreas da imagem que não são necessárias para análise multivariada - e em seguida, com o uso da ferramenta de redução da resolução espacial dos dados (*spatial binning*) - diminuindo-se a resolução da imagem em 5 vezes (altera-se a escala relativa aos pixels). Isso foi necessário devido à grande resolução inicial dos dados, o que gerava imagens que ocupavam muita memória do computador (cada imagem ocupava mais de 1Gb). A Figura 14 demonstra a aplicação das ferramentas citadas.

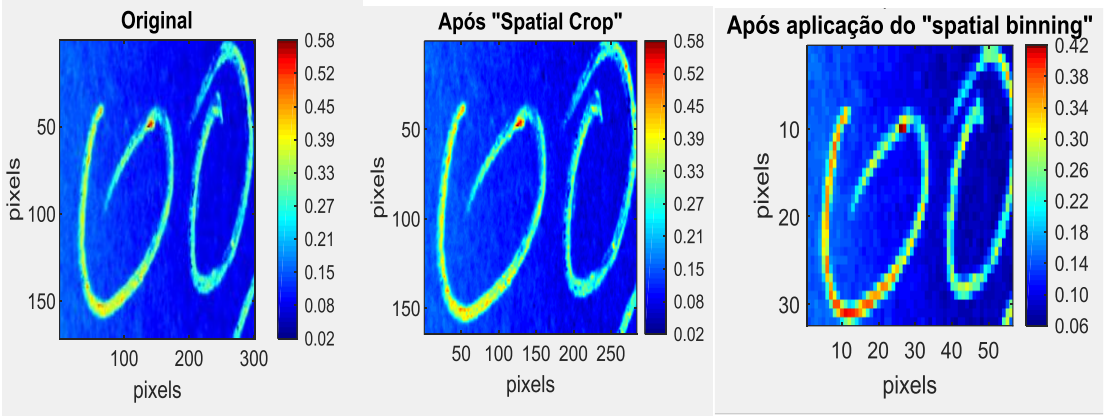


Figura 14: Imagens, em pixels, da amostra 15 de adição de texto, com sequência dos pré-processamentos espaciais aplicados.

Após a seleção da imagem de interesse que contenha os dois zeros relativo às duas tintas, prossegue-se para o pré-processamento dos espectros. O pré-processamento espectral escolhido foi o alisamento por Savitzky-Golay (com janela de 15 pontos e polinômio de grau 2), o qual foi necessário nesse caso devido ao elevado ruído presente nos dados[28]. A figura 15 traz as imagens dos espectros "brutos" e dos espectros "processados", isto é, os espectros após a aplicação do pré-tratamento: percebe-se que os espectros processados foram suavizados com diminuição do ruído.

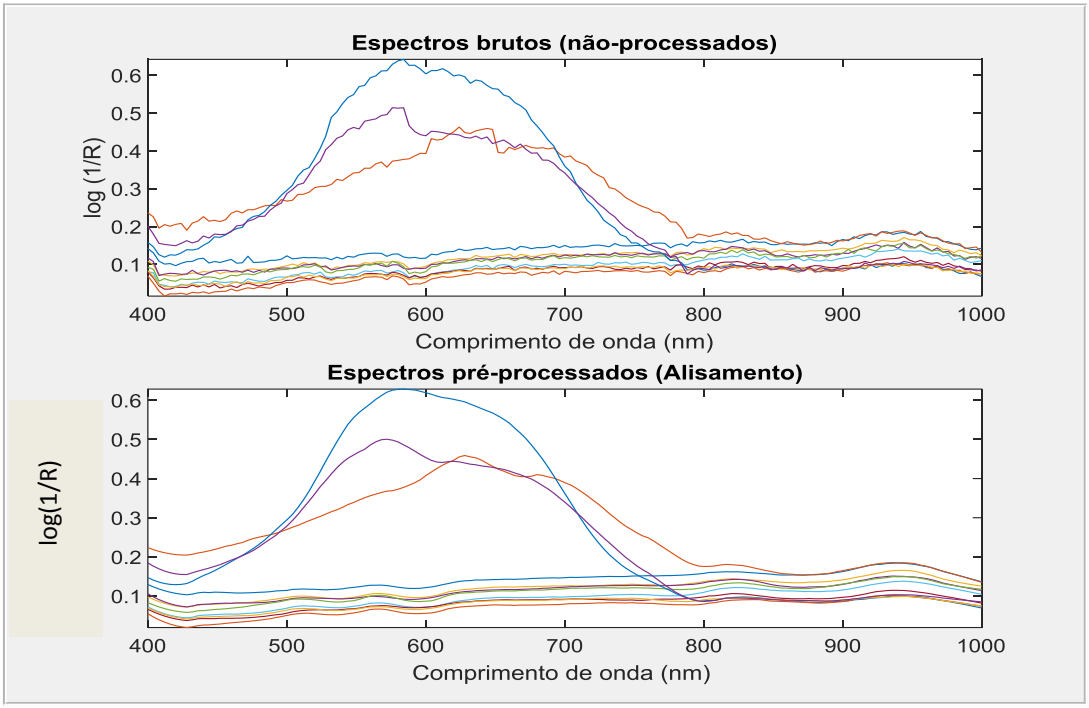


Figura 15: Espectros de vinte pixels antes e após a aplicação do pré-processamento na amostra 15 de adição de texto:(**a**) espectros originais (**b**) com aplicação da técnica de alisamento.

Posteriormente, aplicou-se a ferramenta de seleção de clusters – k means: trata-se de uma máscara com as mesmas dimensões espaciais da imagem hiperespectral e contém as informações dos pixels que serão analisados[27]. Essa máscara é utilizada para selecionar, de forma mais precisa, as regiões da imagem que são de interesse para a análise, no sentido de selecionar apenas os pixels que contém tintas das duas canetas e excluir, portanto, os relativos ao do papel. A figura 16 demonstra a aplicação da ferramenta e traz a imagem separada em 3 "clusters" (agrupamentos): os "clusters" (1) e (2) são relativos às tintas e o "clusters" (3) é o relativo ao papel. Assim, percebe-se que os pixels da região de cor azul e verde (1 e 2) são os referentes às tintas e o de cor

vermelha (3) são os pixels do papel. Não há pixel da cor vermelha na região que não é de tinta de caneta o que demonstra a boa separação decorrente do uso da máscara. Foram selecionados apenas os pixels dos "clusters" 1 e 2, relativos às tintas das canetas. Como para cada pixel, tem-se o espectro correspondente, a aplicação da ferramenta permite a seleção de dados espaciais e espectrais apenas das tintas que são aquelas que são o objeto da análise, a região na qual há a parte adulterada. A aplicação da máscara retira a informação relativa ao papel uma vez que esses pixels podem influenciar muito o modelo PCA por representarem grande parte da variância dos dados.

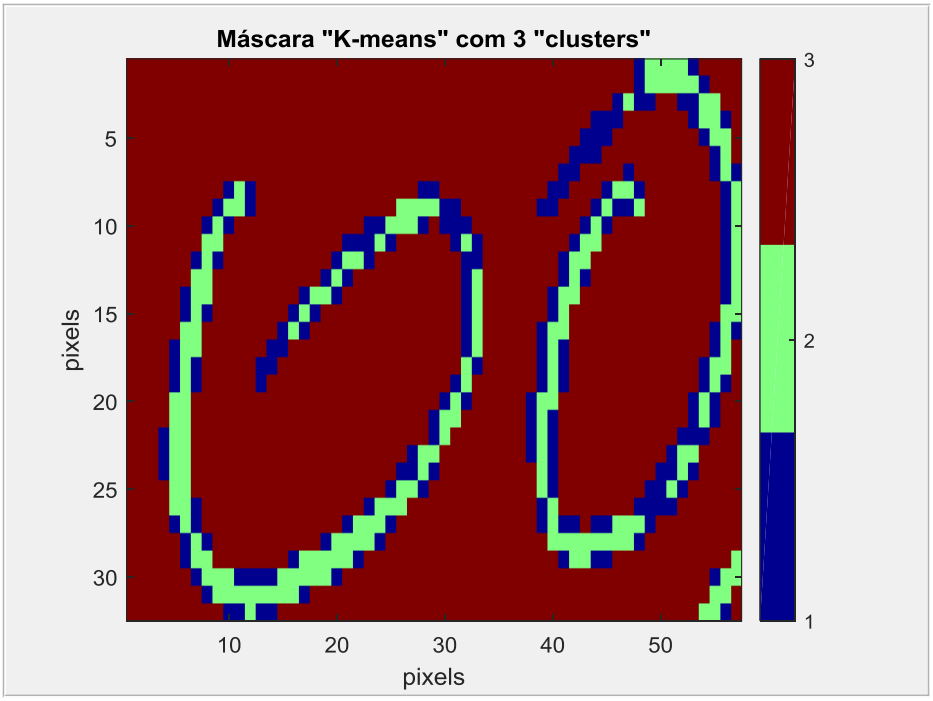
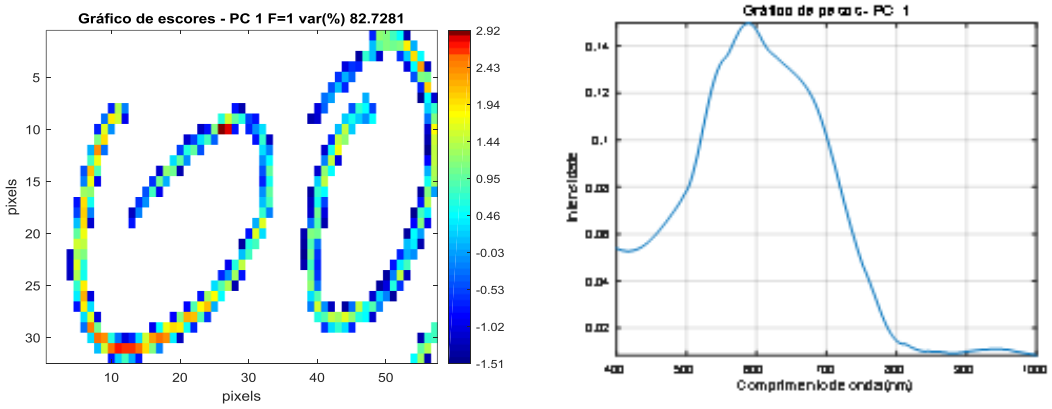


Figura 16: Aplicação da ferramenta "*k-means*" - amostra 15 de adição de texto: foram selecionados os pixels dos *clusters* 1 e 2, referentes às canetas.

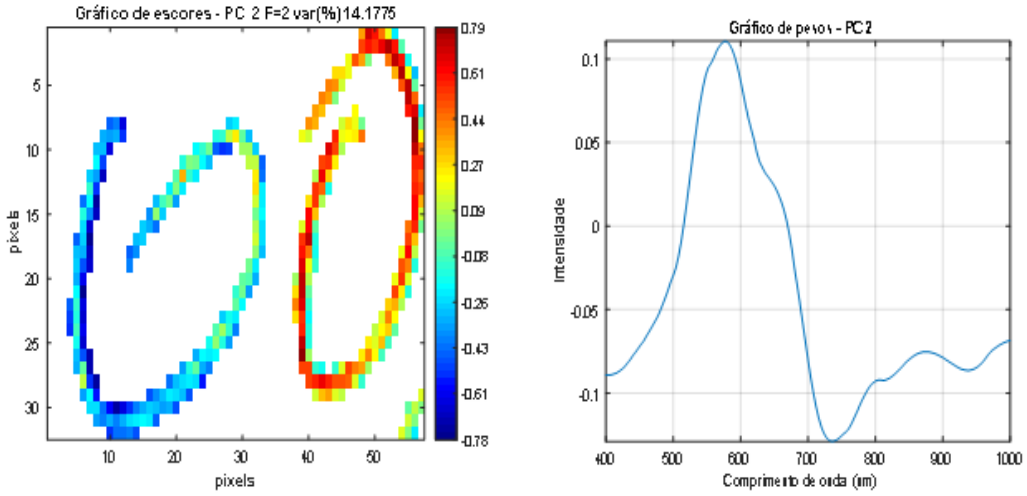
Após o uso da máscara e seleção dos pixels e espectros de interesse, aplica-se a PCA. Para o cálculo das matrizes de escores e de pesos, aplicou-se o algoritmo de Decomposição por Valores Singulares – SVD: o método SVD é a técnica numérica mais acurada e estável para o cálculo das componentes principais[26]. Todos os dados foram centrados na média e utilizou-se 10 componentes principais, que explicaram sempre mais de 98% da variância dos dados encontrados para todas as amostras. A análise por PCA aplicada a dados hiperespectrais fornece como resultado um gráfico de escores, em pixels, que gera uma imagem com as informações dessa PC e um gráfico de pesos, com a importância de cada variável, relativa aos comprimentos de onda. A

análise se demonstra satisfatória para discriminar os dois zeros e, portanto, as duas tintas, quando é possível obter, no gráfico de escores, cores diferentes para zeros diferentes (que correspondem a tintas diferentes). Assim, se a análise tiver sido eficiente para diferenciar o zero adicionado, serão observados pixels de cores/intensidades diferentes para os dois zeros. Os pesos correspondem a uma mensuração das variáveis com maior contribuição para a discriminação no gráfico de escores.

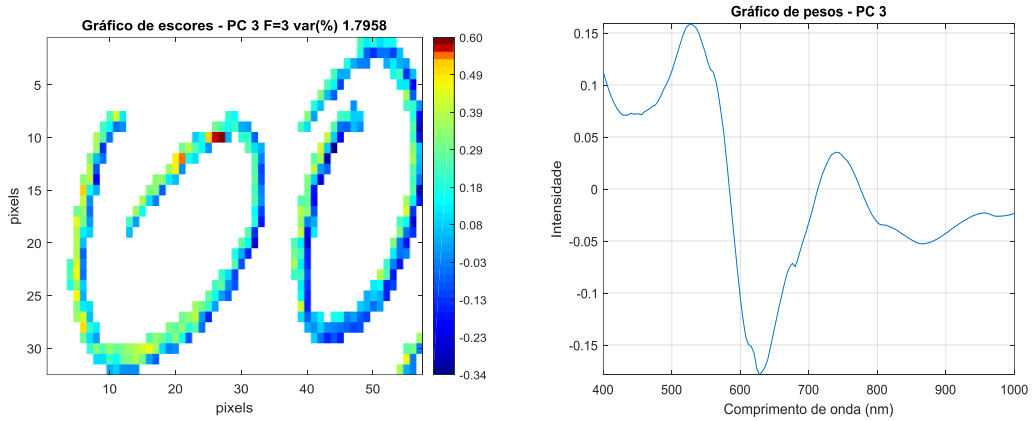
A figura 17 traz, de (**a**) a (**d**), os melhores resultados obtidos de análise por PCA, com os gráficos de escores e os respectivos gráficos de pesos.



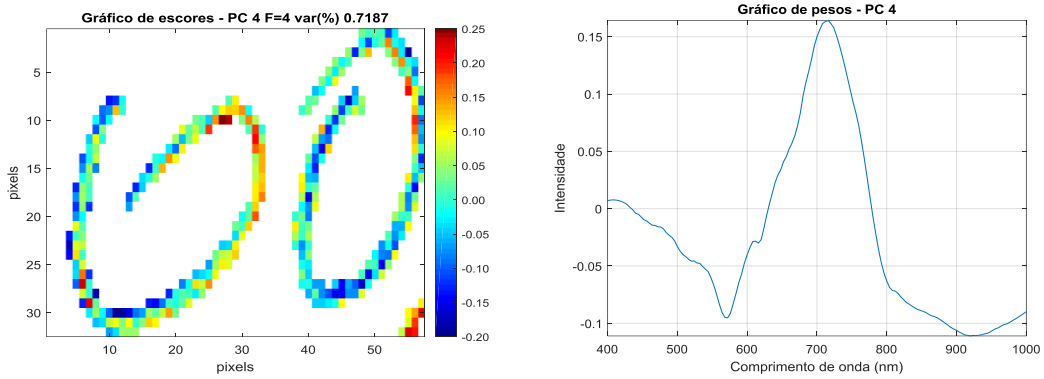
(a) Gráfico de escores (esquerda) e Gráfico de Pesos (direita) para PC 1: Percentual de variância explicada = 83%.



(b) Gráfico de escores (esquerda) e Gráfico de Pesos (direita) para PC 2. Percentual de variância explicada = 14 %.



(c) Gráfico de escores (esquerda) e Gráfico de Pesos (direita) para PC 3. Percentual de variância explicada = 2 %.



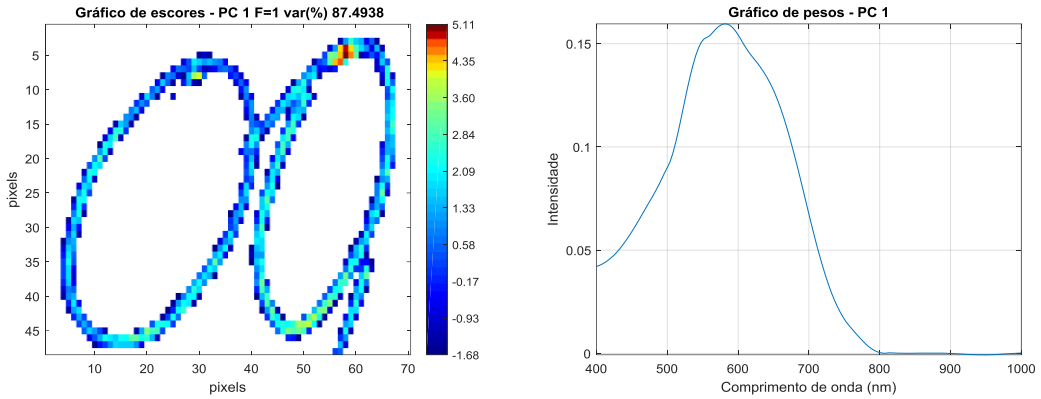
(d) Gráfico de escores (esquerda) e Gráfico de Pesos (direita) para PC 4: Percentual de variância explicada = ~1%

Figura 17: Imagem dos escores das 4 primeiras PCS realizadas para a amostra 15 de adição de texto e os respectivos gráficos de pesos: **(a) PC1; (b) PC 2;(c) PC3 e (d) PC4.**

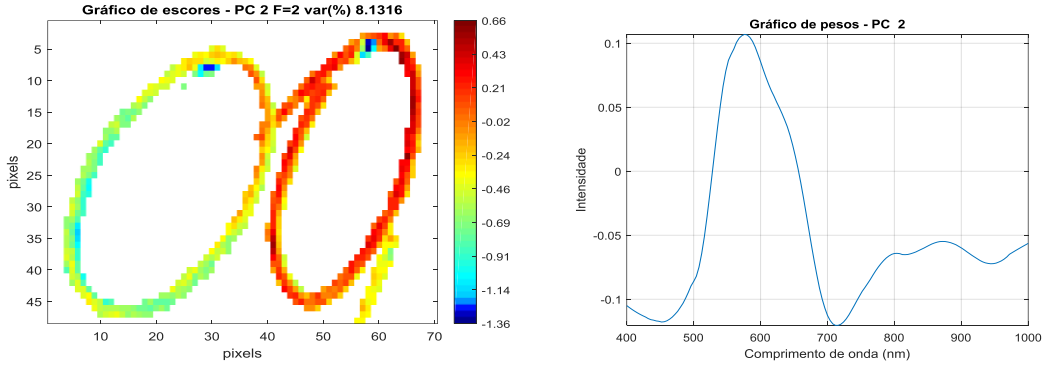
Foi possível explicar, com o uso de 4 componentes principais, mais de 98% da variância dos dados. Observa-se, na figura 17b, que o gráfico de escores da PC 2 apresenta intensidades dos escores nitidamente diferentes para os pixels dos dois zeros: o segundo zero está com um conjunto de tons próximos à laranja e vermelho e o primeiro zero com tons próximos a azul e verde. Isso demonstra que a técnica de PCA foi capaz de diferenciar os pixels e os espectros da tinta 1 e da tinta 2, discriminando-as. O gráfico de pesos demonstra que os comprimentos de onda mais importantes para a discriminação entre as tintas estão próximos à faixa de 600 nm.

Outro exemplo em que é possível se verificar a aplicação da técnica de PCA é a amostra de número 35 na qual o primeiro “0”, identificado como tinta 1, foi escrito com a caneta 13 e o zero adicionado, identificado como tinta 2, foi escrito com a caneta 15.

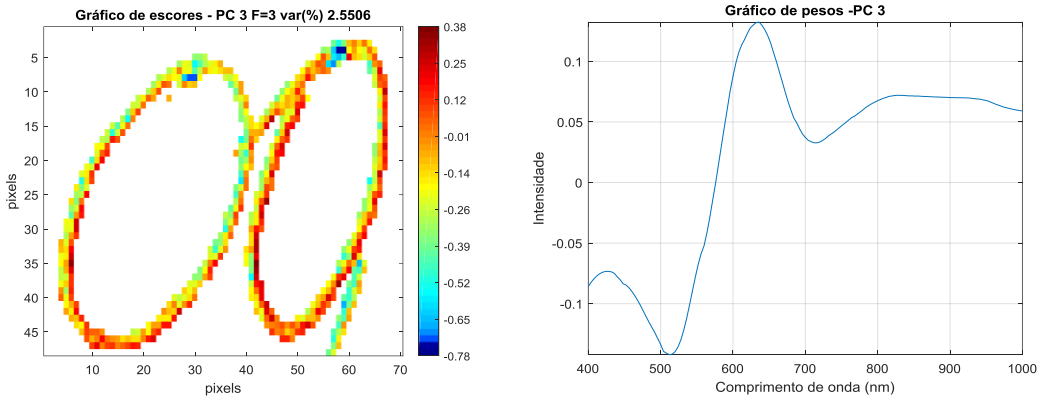
A figura 18, traz os gráficos de escores e de pesos, de (**a**) a (**d**), com os melhores resultados obtidos, com a PCA das imagens de escores e os respectivos gráficos de pesos.



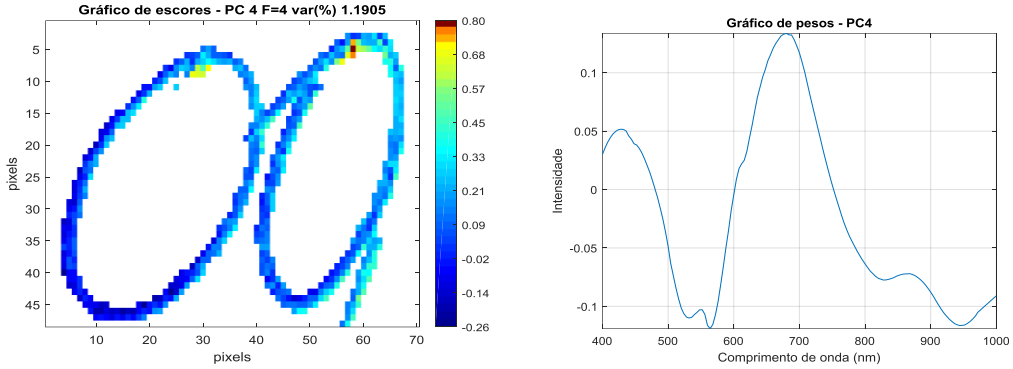
(a) Gráfico de escores (esquerda) e Gráfico de Pesos (direita) para PC 1: Percentual de variância explicada = 88%.



(b) Gráfico de escores (esquerda) e Gráfico de Pesos (direita) para PC 2. Percentual de variância explicada = 8 %.



(c) Gráfico de escores (esquerda) e Gráfico de Pesos (direita) para PC 3. Percentual de variância explicada = 3%.



(d) Gráfico de escores (esquerda) e Gráfico de Pesos (direita) para PC 4. Percentual de variância explicada = 1 %.

Figura 18: Imagem dos escores das 4 primeiras PCS realizadas para a amostra 35 de adição de texto e os respectivos gráficos de pesos: **(a) PC1; (b) PC 2;(c) PC3 e d) PC4.**

Conforme se observa na figura 18b, no gráfico de escores da PC 2, há intensidades de escores diferentes para os dois zeros demonstrando que são de tintas diferentes. A análise do gráfico de pesos da PC2 demonstra que a faixa espectral responsável pela maior discriminação é a próxima à região de 600 nm.

Ao todo, a análise por PCA demonstrou-se satisfatória para 18 amostras de adição de texto, representando 36% do total de amostras produzidas.

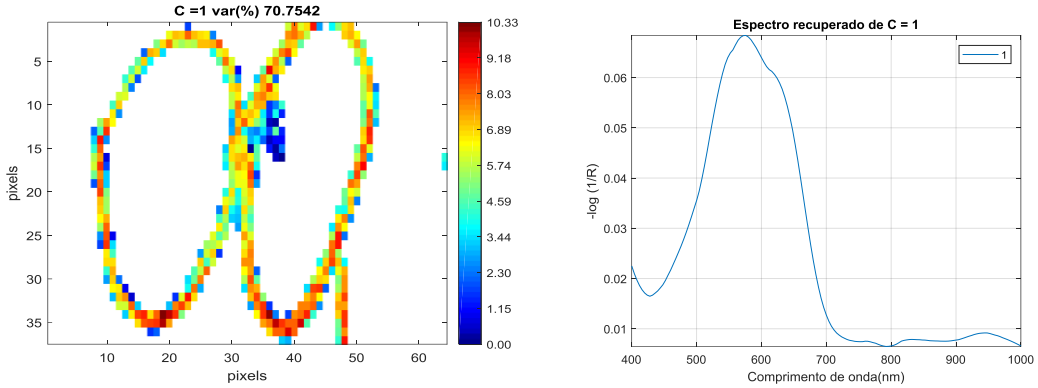
- **Análise Multivariada de Curvas – MCR**

A amostra tomada como exemplo foi a amostra número 21 de adição de texto, na qual o primeiro “0”, identificado como tinta 1, foi escrito com a caneta 1 e o zero adicionado, identificado como tinta 2, foi escrito com a caneta 10.

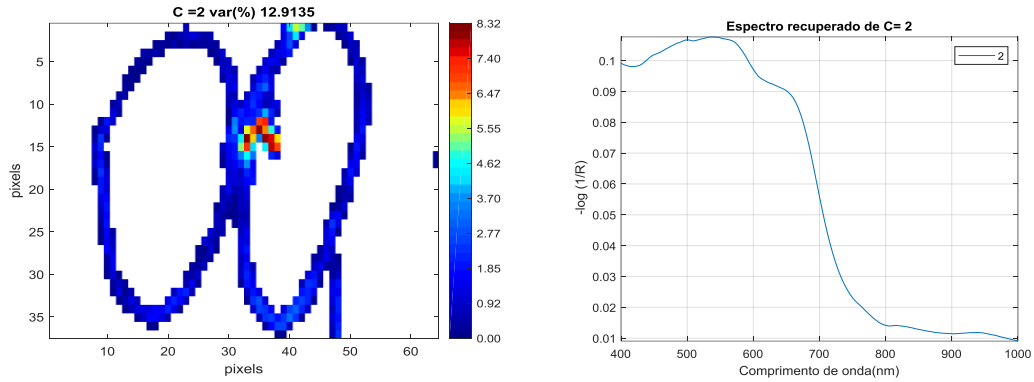
O tratamento dos dados é semelhante ao já descrito com a técnica de análise por PCA. Aplicam-se os mesmos tratamentos de redução do tamanho da imagem ("spatial cropping" e "spatial binning"), bem como o mesmo pré-processamento para suavização dos espectros ("Smoothing Sav Gol", com janela de 15 e polinômio de grau 2). A seleção dos pixels e dos espectros das tintas 1 e 2 é realizada pela aplicação da máscara k-means e posteriormente aplica-se a técnica de PCA. Quando a técnica de PCA, pelo gráfico de escores e pesos, não se mostra suficiente para diferenciar os zeros, procede-se, então, a análise por MCR. A análise por MCR foi feita com a seleção das componentes pela técnica "PURITY" e com a aplicação da restrição de não-negatividade para concentração. A técnica "PURITY" é um processo de otimização que estima os valores das intensidades dos componentes puros presentes na matriz de dados, selecionando as colunas com as variáveis mais puras de acordo com o número de fatores que se acredita existirem na amostra[23].O objetivo da aplicação de restrição é diminuir a possibilidade de resultados diferentes para um mesmo conjunto de dados. A restrição de não-negatividade impõe que todas as soluções sejam maiores ou iguais a zero. Esse tipo de restrição aplica-se a todos os perfis de concentração[22].

A interpretação dos resultados da análise por MCR também ocorre pela identificação de valores de intensidades relativas (equivalentes aos escores de PCA) distintas os zeros referentes às duas tintas: a técnica terá discriminado os zeros e, portanto, as duas tintas se forem observadas, no mapa de distribuição, cores diferentes para os zeros. O gráfico da estimativa do espectro puro indica a região da faixa espectral responsável pela maior discriminação entre as tintas.

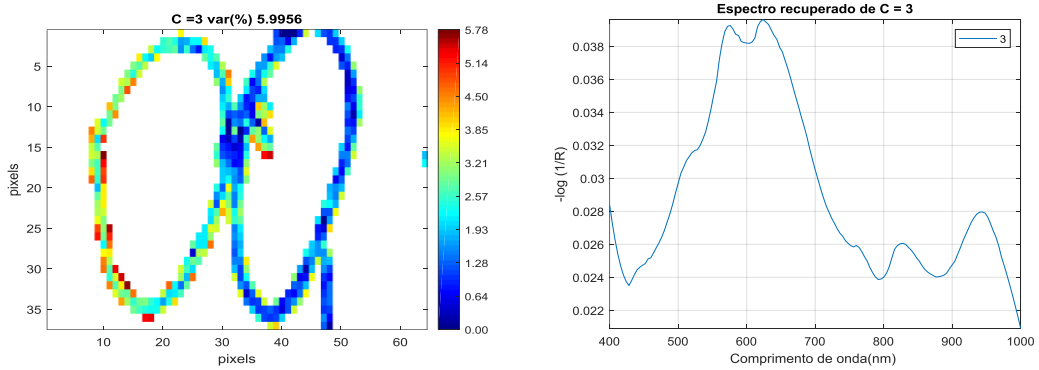
As análises por MCR foram testadas com até 6 componentes da mistura. Os resultados eram analisados e selecionavam-se as imagens das componentes que melhor discriminassem os zeros. A amostra retratada como exemplo foi a de número 21 de adição de texto, a análise foi feita com 3 componentes, as quais explicaram, no total, cerca de 89 % da variância dos dados, e os resultados podem ser observados na Figura 19, que traz os mapas de distribuição e de espectros recuperados, de (**a**) a (**c**).



(a) Mapas de distribuição (esquerda) e Espectro puro (direita) para C=1: Percentual de variância explicada = 71%.



(b) Mapas de distribuição (esquerda) e Espectro puro (direita) para C= 2. Percentual de variância explicada = 13 %.



(c) Mapas de distribuição (esquerda) e Espectro puro (direita) para C= 3. Percentual de variância explicada = 6 %.

Figura 19: Imagem dos mapas de distribuição das 3 primeiras componentes realizadas para a amostra número 21 de adição de texto e os respectivos gráficos de espectros puros: **(a) C=1; (b) C=2; (c) C=3.**

Assim, observa-se no mapa de distribuição, no item **(c)** na figura 19, a diferenciação entre tons de cores dos dois zeros e a distinção, portanto, entre o número que foi escrito com a caneta 1 do que foi escrito com a caneta 10. Pelo gráfico do espectro recuperado, percebe-se que a região de faixa espectral responsável pela maior diferenciação entre as tintas das canetas 1 e 10 é a compreendida entre os comprimentos de onda de 600 a 700 nm.

Para fins de comparação, a figura 20 traz os gráficos de escores obtidos das análises por PCA, demonstrando que a análise por MCR permite melhor discriminação entre as duas tintas.

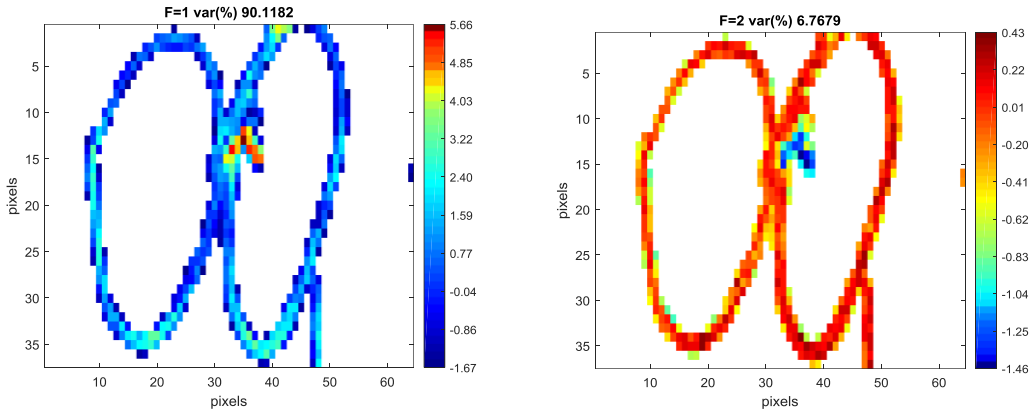
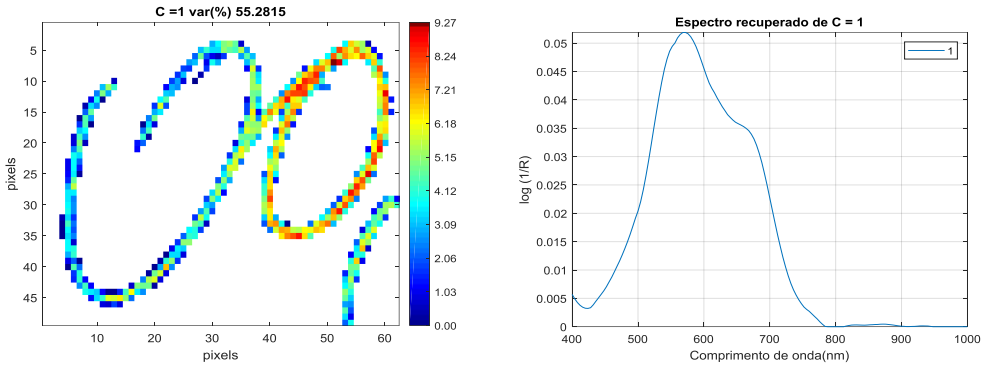


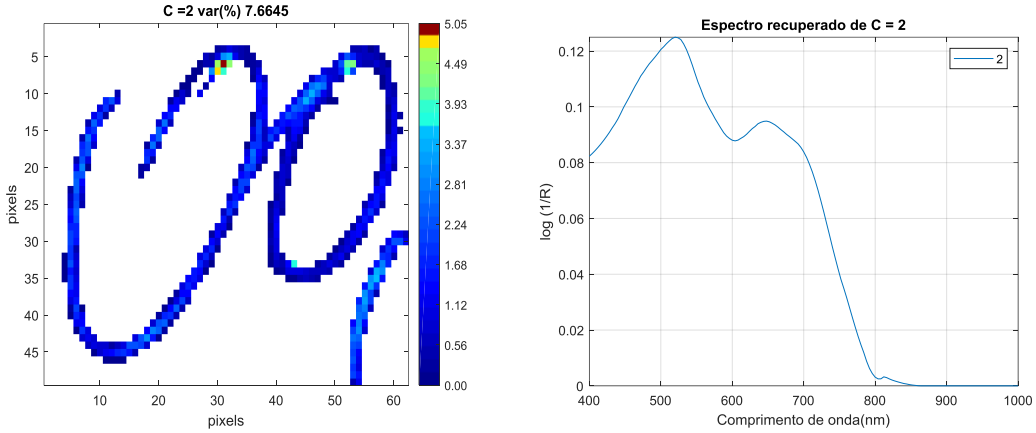
Figura 20: Gráfico de escores obtidos pelas análises de PCA, para as duas primeiras componentes principais, da amostra número 21 de adição de texto.

Outro exemplo em que é possível observar a aplicação da técnica de MCR foi a amostra 42 de adição de texto, na qual o primeiro “0”, identificado como tinta 1, foi escrito com a caneta 13 e o zero adicionado, identificado como tinta 2, foi escrito com a caneta 12.

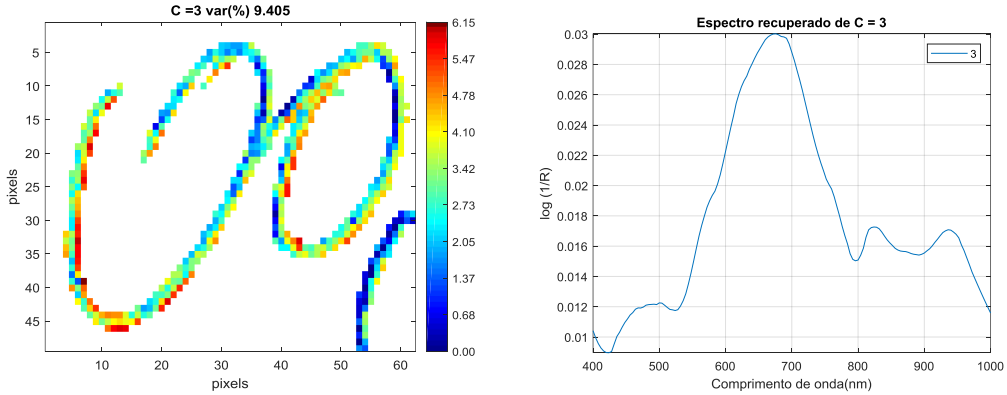
A análise dessa amostra foi feita com 4 componentes, explicando-se um total de 79% da variância dos dados, e a figura 21 traz os melhores resultados.



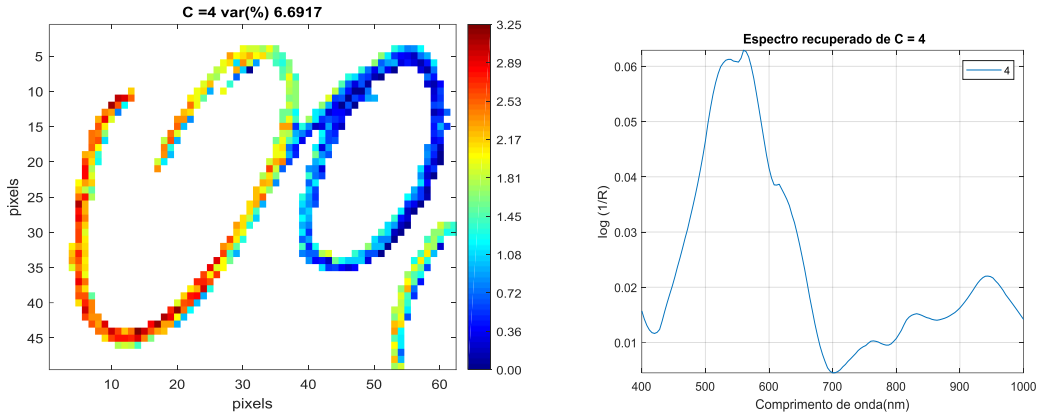
(a) Mapas de distribuição (esquerda) e Espectro puro (direita) para C=1: Percentual de variância explicada =55 %.



(b) Mapas de distribuição (esquerda) e Espectro puro (direita) para C=2. Percentual de variância explicada = 8 %.



(c) Mapas de distribuição (esquerda) e Espectro puro (direita) para C=3. Percentual de variância explicada = 9%.



(d) Mapas de distribuição (esquerda) e Espectro puro (direita) para C=4. Percentual de variância explicada = 7 %.

Figura 21: Imagem dos mapas de distribuição das 4 componentes selecionadas para a amostra 42 de adição de texto e os respectivos gráficos de espectros: **(a) C=1; (b) C=2; (c) C=3 e (d) C=4.**

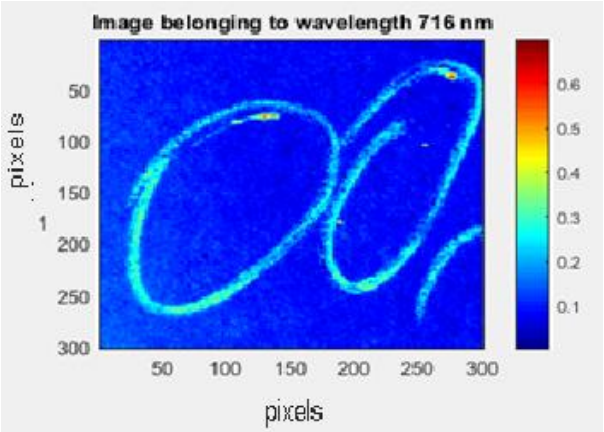
Conforme se observa na figura 21, na Componente 4, foi visível a discriminação entre os tons de cores dos dois zeros e percebe-se a diferenciação do trecho que foi escrito com a caneta 13 do que foi escrito com a caneta 12. Pela análise do espectro recuperado, observa-se que a região de faixa espectral responsável pela maior diferenciação entre as tintas das canetas 13 e 12 é a compreendida entre 500 a 600 nm.

Ao todo, a análise por MCR demonstrou-se satisfatória para 3 amostras de adição de texto, representando 6% do total de amostras produzidas.

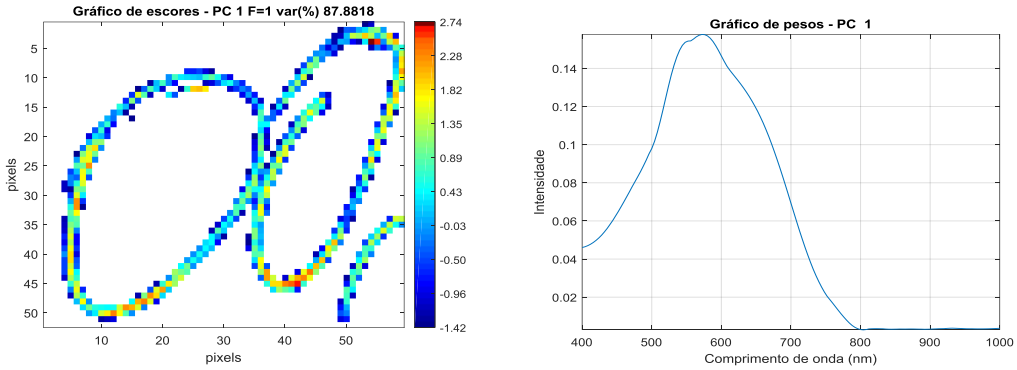
- **Inconclusivo**

A amostra tomada como exemplo foi a amostra número 40 de adição de texto, na qual o primeiro “0”, identificado como tinta 1, foi escrito com a caneta 13 e o zero adicionado, identificado como tinta 2, foi escrito com a caneta 17.

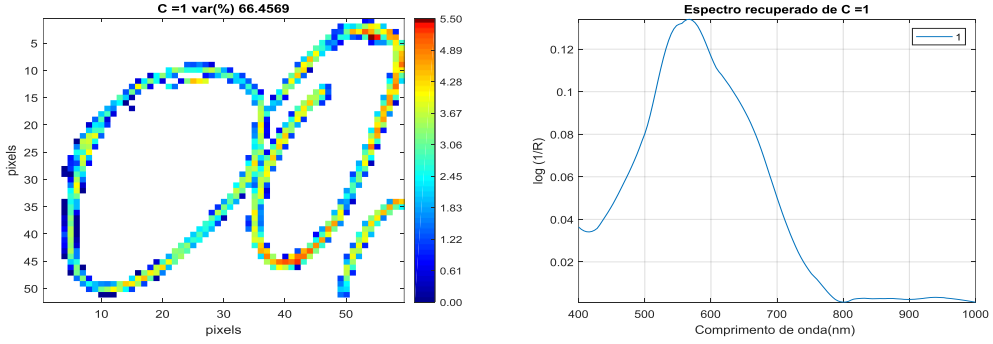
O resultado foi considerado inconclusivo porque após a aplicação o tratamento dos dados, com os pré-processamentos espaciais e espectrais já descritos nas seções anteriores e a utilização das técnicas de análise univariada e multivariada – por PCA e por MCR – não foi possível diferenciar os dois zeros. A Figura 22 traz as imagens (de **a** a **c**), com os melhores resultados obtidos.



(a) Exemplo de imagem, em pixel, correspondente a comprimento de onda de 716 nm (não foi possível obter discriminação para nenhum dos comprimentos de onda)



(b) Exemplo de gráfico de escores (esquerda) e Gráfico de Pesos (direita) para PC=1: Percentual de variância explicada = 88%.



(c) Exemplo de mapas de distribuição (esquerda) e Espectro puro (direita) para C=1: Percentual de variância explicada =66%.

Figura 22: Exemplos de análises inconclusivas para amostra número 40 de adição de texto: (**a) Univariada (b) Multivariada - PCA (c) Multivariada – MCR.**

Uma possível explicação para a não discriminação entre os zeros pode ser a semelhança na composição das tintas das canetas analisadas.

Ao todo, as análises demonstraram-se inconclusivas para 4 amostras de adição de texto, representando 8% do total de amostras produzidas.

- **Resumo das Análises de Adição de Texto**

Foram analisadas 50 amostras de adição de texto, para as quais o método de análise univariada foi eficaz para 50 % dos casos e o de análise multivariada, para 42% dos casos, sendo 36% com a utilização da de PCA e 6% com o uso de MCR. Para um total de 8% dos casos (4 amostras), não foi possível diferenciar os dois zeros, gerando um resultado inconclusivo para a detecção da falsificação do documento por adição de um texto. Uma explicação para a não impossibilidade de discriminação entre os zeros das tintas pode ser a semelhança na composição das tintas das canetas. A tabela número 7 traz uma compilação de todos os resultados obtidos para as amostras de adição de texto.

Tabela 7. Total de resultados das Análises - Adição de Texto

Número da amostra	Canetas utilizadas	Método*	Número da amostra	Canetas utilizadas	Método*
1	16 e 1	Uni	26	1 e 12	PCA
2	16 e 13	PCA	27	1 e 2	Uni
3	16 e 6	PCA	28	9 e 7	Uni
4	16 e 15	PCA	29	9 e 10	Uni
5	16 e 7	PCA	30	9 e 8	Uni
6	16 e 4	Uni	31	9 e 14	Uni
7	16 e 10	Uni	32	9 e 17	Uni
8	16 e 11	Inconclusivo	33	9 e 2	Uni
9	16 e 8	PCA	34	13 e 9	Uni
10	16 e 14	PCA	35	13 e 15	PCA
11	16 e 17	PCA	36	13 e 7	PCA
12	16 e 3	PCA	37	13 e 10	Uni
13	16 e 12	PCA	38	13 e 8	PCA
14	16 e 5	PCA	39	13 e 14	Inconclusivo
15	16 e 2	PCA	40	13 e 17	Inconclusivo
16	1 e 13	Uni	41	13 e 3	MCR
17	1 e 6	Uni	42	13 e 12	MCR
18	1 e 15	Uni	43	13 e 2	PCA
19	1 e 7	Uni	44	10 e 6	Uni
20	1 e 4	Uni	45	10 e 15	Uni
21	1 e 10	MCR	46	10 e 7	Uni
22	1 e 11	Uni	47	10 e 11	Uni
23	1 e 8	Uni	48	6 e 14	PCA
24	1 e 14	Uni	49	6 e 17	Inconclusivo
25	1 e 3	Uni	50	6 e 3	PCA
Total		**Uni (25)**		**50%**	
		PCA (18)		**36%**	
		MCR (3)		**6%**	
		Inconclusivo (4)		**8%**	

*****Legenda: Uni=univariado; PCA= Análise de Componentes Principais e MCR= Resolução Multivariada de Curvas.

5.2. Obliteração

Conforme descrito, os resultados obtidos podem ser descritos em função da técnica necessária para a revelação das letras da palavra obliterada. Para a demonstração

representativa das 15 amostras analisadas, será feita, para cada técnica de análise: uma descrição mais detalhada do procedimento utilizado e dos resultados obtidos, para uma amostra; uma descrição mais objetiva, focada nos resultados; para outra amostra. Ao fim, será apresentada uma tabela final, com a compilação dos resultados de todas as amostras analisadas.

- **Análises Univariada**

A amostra tomada como exemplo foi a amostra número 10 de obliteração, na qual a palavra "FALSO" foi escrita com a tinta 1 (caneta 11) e obliterada com a tinta 2 (caneta 12). Após a aquisição dos dados do VSC®, esses são importados para o ambiente Matlab, organizados no formato do cubo hiperespectral e analisados por meio do programa HYPERTools.

Se for possível discriminar as tintas, com a seleção de apenas um comprimento de onda, a imagem permitirá ver o texto da caneta 1 que foi escondido com a caneta 2. A figura 23 demonstra um exemplo de análise em que foi possível revelar as letras escritas que estavam ocultas pela sobrecarga da outra caneta: a figura corresponde à imagem para o comprimento de onda 763 nm. O resultado demonstra que foi possível encontrar uma imagem selecionando-se um comprimento de onda (763 nm) capaz de diferenciar a tinta 1 da tinta 2. Ou seja, esse comprimento de onda é seletivo para uma das tintas de maneira a diferenciá-las na imagem. A resposta é a figura da imagem hiperespectral no comprimento de onda selecionado. A escala de cores na figura 23, em pixels, corresponde à intensidade do espectro obtido em pseudo-absorbância e é crescente da cor azul para a vermelha.

Após essa primeira análise, com a verificação de que a análise univariada é suficiente para separar a mistura de tintas, plotou-se um gráfico em que é possível observar espectros claramente distintos e separados para a tinta 1, para a tinta 2 e para o papel, conforme demonstra a figura 23.

Os gráficos dos espectros médios foram formados a partir da seleção da matriz de dados desdobradas do cubo hiperespectral, no ambiente Matlab, com a seleção dos pixels relativos à caneta 1, à caneta 2 e ao papel.

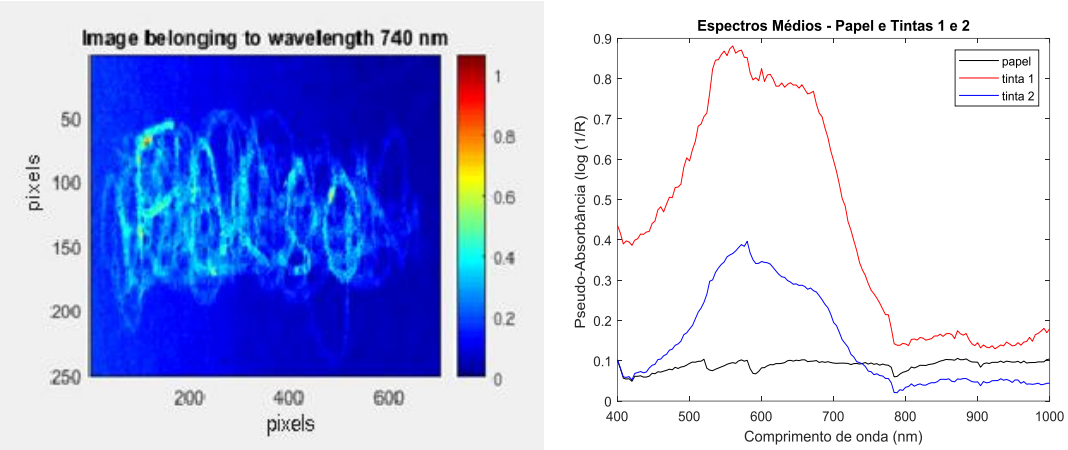


Figura 23 – À esquerda (a), imagem, em pixels, correspondente ao comprimento de onda 740 nm e à direita (b), Gráfico dos Espectros médios de pixels do papel, da tinta 1 e da tinta 2 – amostra 10 de obliteração.

Conforme se observa na Figura 23b, o espectro da tinta 1, em vermelho, é distinto do espectro da tinta 2, em azul, demonstrando que os espectros estão visualmente separados em 740 nm sem a necessidade de utilização de nenhum tratamento estatístico adicional.

Outro exemplo em que é possível verificar-se a revelação da palavra escrita é o da amostra número 12 de obliteração, na qual a palavra "FALSO" foi escrita com a tinta 1 (caneta 11) e obliterada com a tinta 2 (caneta 17). A figura 24 demonstra que foi possível revelar as letras escritas que estavam ocultas pela sobrecarga da outra caneta com a imagem selecionada para o comprimento de onda 750 nm.

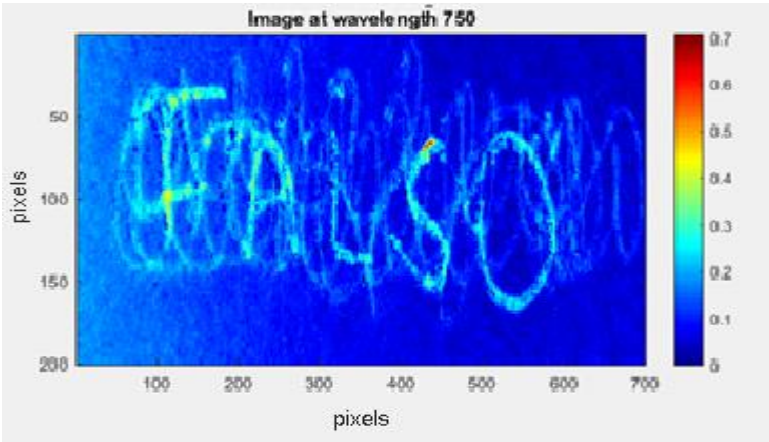


Figura 24 – Imagem, em pixels, correspondente ao comprimento de onda 750 nm - amostra 12 de obliteração.

Ao todo, foi possível, por meio da análise univariada, identificar a palavra "FALSO", em 3 amostras, de um total de 15, com todas as letras identificadas, representando 20% do total de amostras produzidas.

- **Análise Multivariada**

  Análise de Componentes Principais – PCA

A amostra tomada como exemplo foi a amostra número 9 de obliteração, na qual a palavra "FALSO" foi escrita com a tinta 1 (caneta 5) e obliterada com a tinta 2 (caneta 17).

De acordo com a metodologia proposta, inicialmente procede-se à análise univariada, em busca de um comprimento de onda que seja seletivo para uma das tintas, e uma vez que a técnica não se demonstre eficaz, procede-se à análise multivariada, começando-se por PCA. A figura 25 demonstra a imagem obtida para a análise univariada da amostra e percebe-se que não há distinção visível dos pixels capaz de revelar nenhum caractere da palavra escondida. O equipamento realiza a análise ao longo da faixa espectral de 400 a 1000 nm, trazendo várias imagens. Entretanto, nenhuma delas revelou algum caractere do texto obliterado. A imagem correspondente à figura 25 demonstra um exemplo de uma das imagens obtidas na varredura.

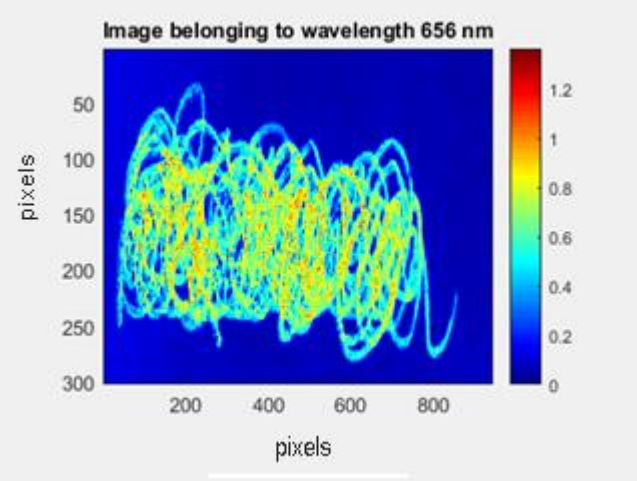


Figura 25 – Imagem, em pixels, correspondente ao comprimento de onda 656 nm- amostra 9 de obliteração.

Na continuidade da análise, os dados são pré-processados para se realizar a análise multivariada: utilizaram-se as ferramentas de pré-tratamento espacial: ferramenta “spatial cropping”, excluindo-se as áreas da imagem que não serão necessárias para análise multivariada, reduzindo o tamanho dos dados a serem analisados - e, em seguida, a aplicação da ferramenta “spatial binning”- tornando a resolução da imagem 5 vezes menor. A figura 26 demonstra a aplicação das ferramentas citadas.

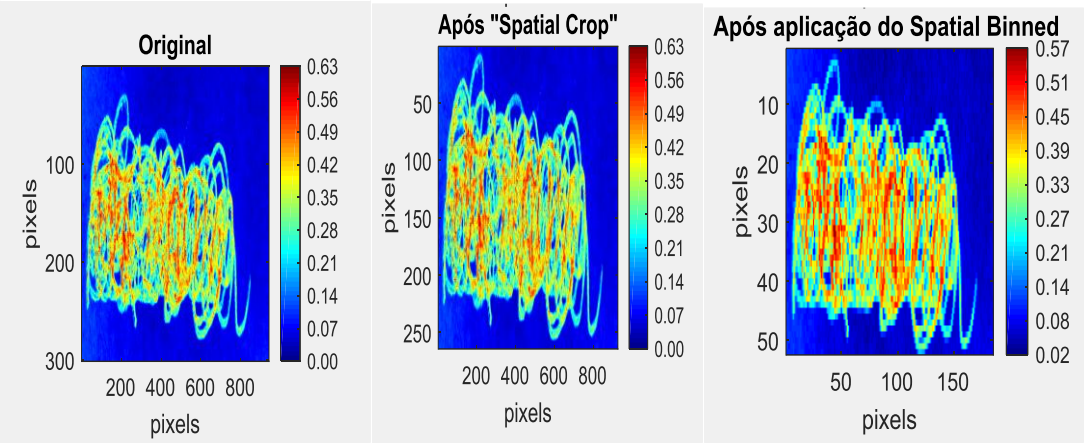


Figura 26: Imagens, em pixels, da amostra 9 de obliteração, com sequência dos pré-processamentos espaciais aplicados.

Com a seleção da imagem de interesse que contenha as duas tintas, prossegue-se para o pré-processamento dos espectros. O pré-processamento espectral aplicado foi o alisamento por Savitzky-Golay (com janela de 15 pontos e polinômio de grau 2). A figura 27 ilustra o resultado obtido, com os espectros “brutos” e os espectros “pré-processados”: percebe-se que os espectros pré-processados foram suavizados com diminuição do ruído.

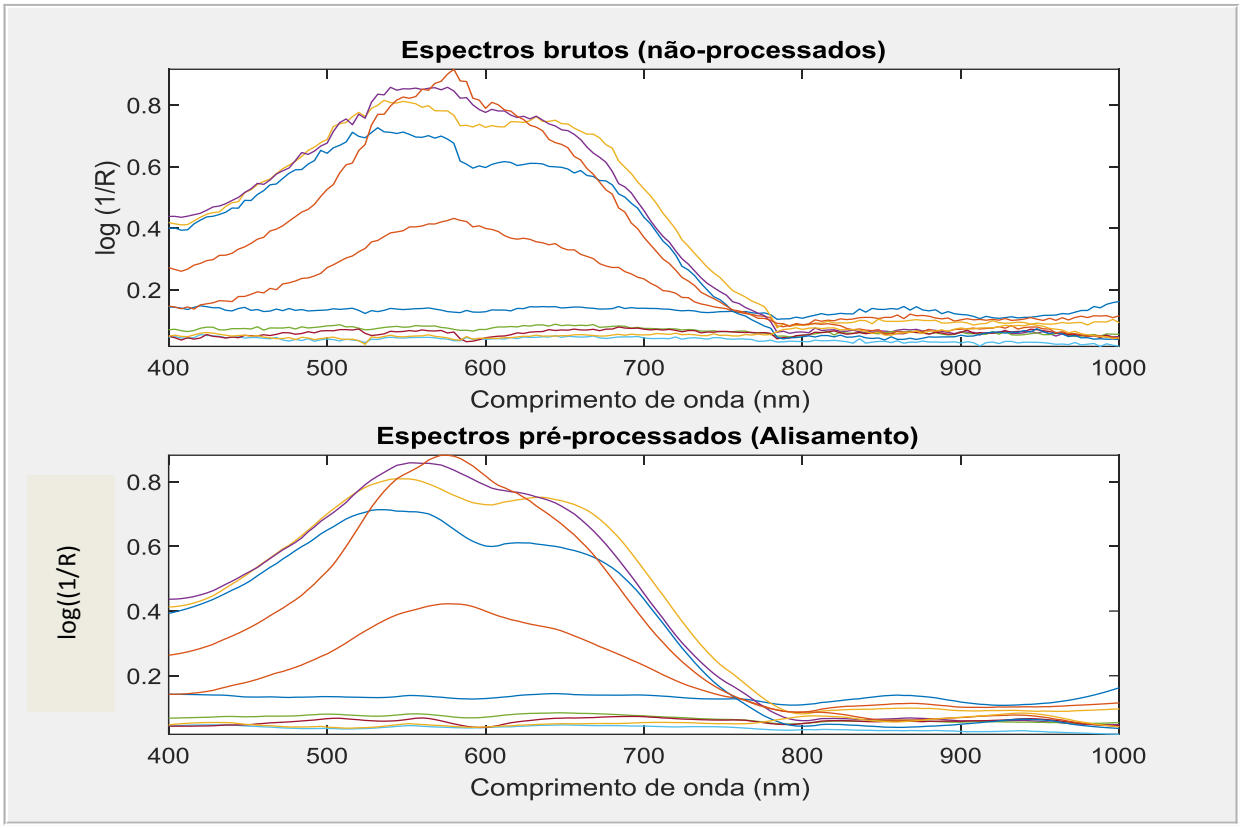


Figura 27: Espectros de vinte pixels antes e após a aplicação do pré-processamento na amostra 9 de obliteração:(**a**) espectros originais (**b**) com aplicação da técnica de alisamento.

Posteriormente, aplicou-se a ferramenta de seleção de clusters pela máscara k means, para excluir os pixels relativos ao do papel. A figura 28 demonstra a aplicação da ferramenta, com a separação da imagem em 3 “clusters”: (1 e 3) são as tintas e (2) é o papel. Os pixels da região de cor vermelha e azul (1 e 2) são os relativos às tintas e o de cor verde (3) são os relativos ao papel. Não se enxergam pixels da cor vermelha ou azul na região que não é de tinta de caneta e isso demonstra a boa separação decorrente do uso da máscara. Foram selecionados os pixels dos “clusters” (1) e (3) porque são os relativos às tintas das canetas. A aplicação da máscara faz com que o papel não seja analisado porque ele não é de interesse para a revelação dos caracteres escondidos.

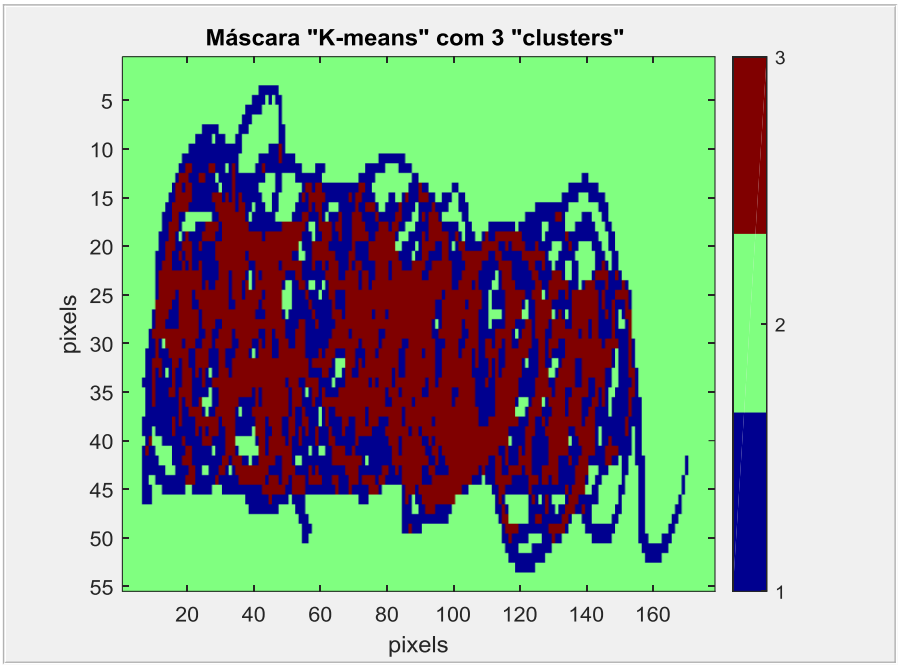
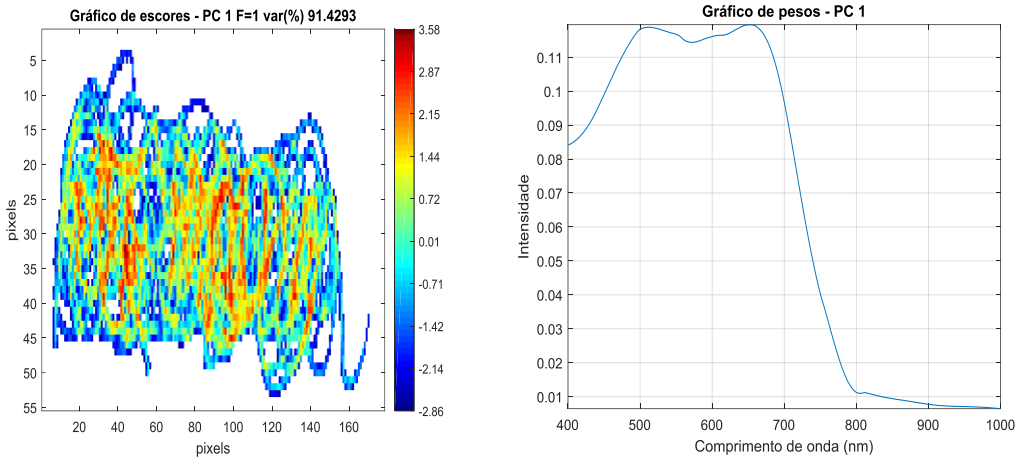


Figura 28: Aplicação da ferramenta "*k-means*" - amostra 9 de obliteração: foram selecionados os pixels dos *clusters* 1 e 3, referentes às canetas.

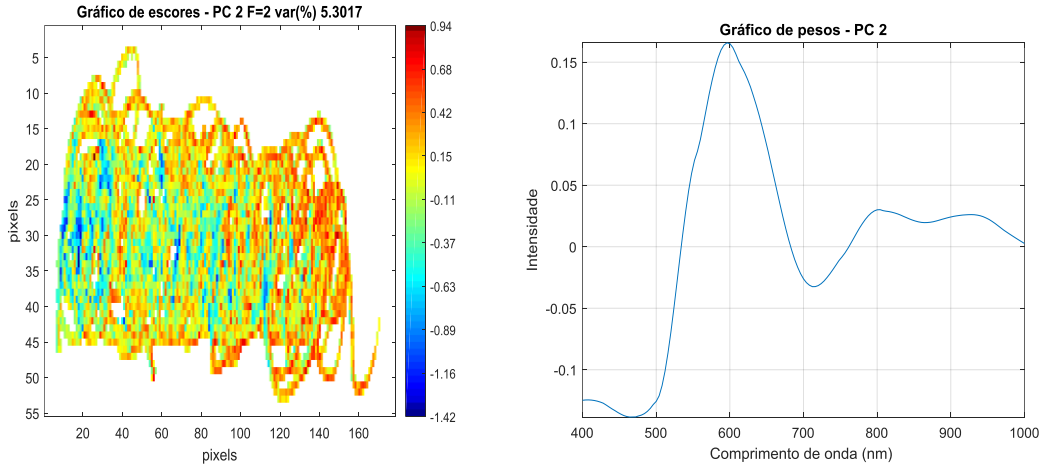
A ferramenta de seleção de *clusters* permite a seleção dos pixels e dos espectros relativos apenas às tintas das canetas

Após o uso da máscara, aplicou-se a técnica PCA. Para o cálculo das matrizes de escores e de pesos, utilizou-se o algoritmo SVD. Todos os dados foram centrados na média e utilizou-se 10 componentes principais, as quais explicam mais de 98% da variância dos dados encontrados. Como anteriormente, a análise por PCA se demonstra satisfatória para separar a mistura das duas tintas quando é possível obter intensidades diferentes para tintas diferentes, no gráfico de escores. Assim, se a análise tive sido eficiente para revelar as letras, serão observados pixels de uma mesma tonalidade, com uma sequência que forme a letra revelada. Esses pixels estarão com uma cor distinta do restante da imagem, que é correspondente à outra tinta. Os pesos mensuram as variáveis com maior contribuição para a discriminação das tintas no gráfico de escores.

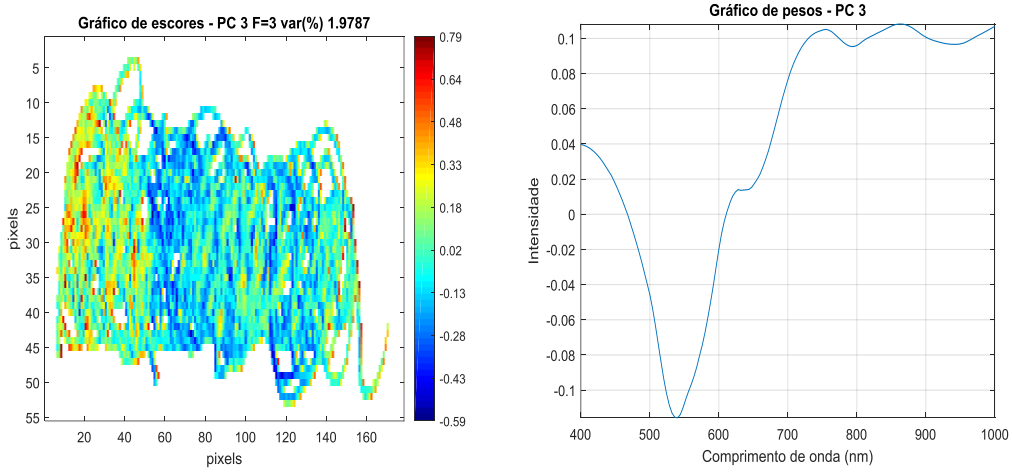
A figura 29 traz, de (**a**) a (**d**), a seleção dos melhores resultados de análise por PCA, com os gráficos de escores e os gráficos de pesos.



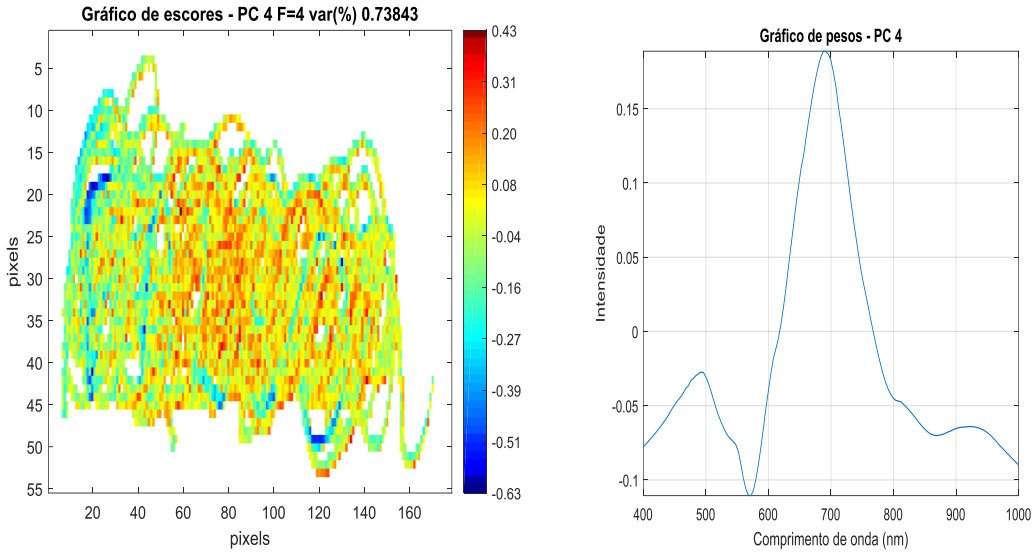
(a) Gráfico de escores (esquerda) e Gráfico de Pesos (direita) para PC 1: Percentual de variância explicada = 91%.



(b) Gráfico de escores (esquerda) e Gráfico de Pesos (direita) para PC 2. Percentual de variância explicada = 5 %.



(c) Gráfico de escores (esquerda) e Gráfico de Pesos (direita) para PC 3. Percentual de variância explicada = 2 %.



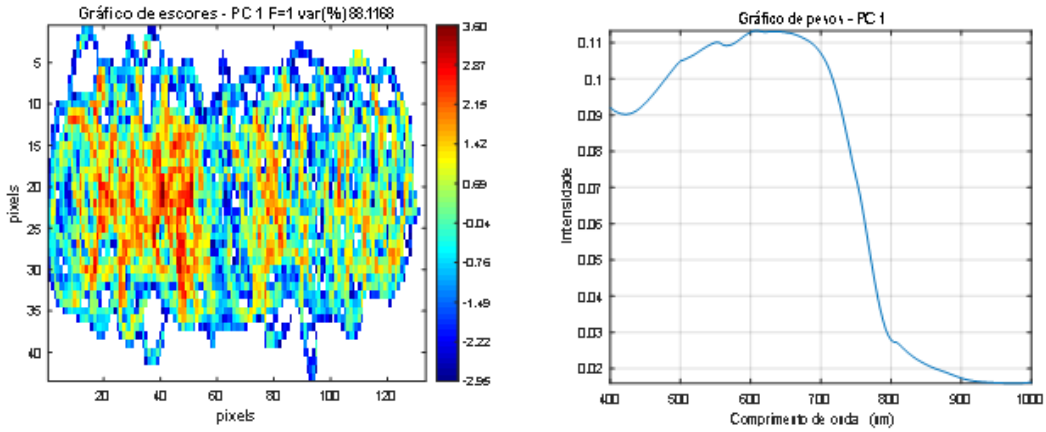
(d) Gráfico de escores (esquerda) e Gráfico de Pesos (direita) para PC 4: Percentual de variância explicada = ~1%

Figura 29: Imagem dos gráficos de escores das 4 primeiras PCS e dos respectivos gráficos de pesos para a amostra 9 de obliteração: **(a) PC1; (b) PC 2;(c) PC3 e (d) PC4.**

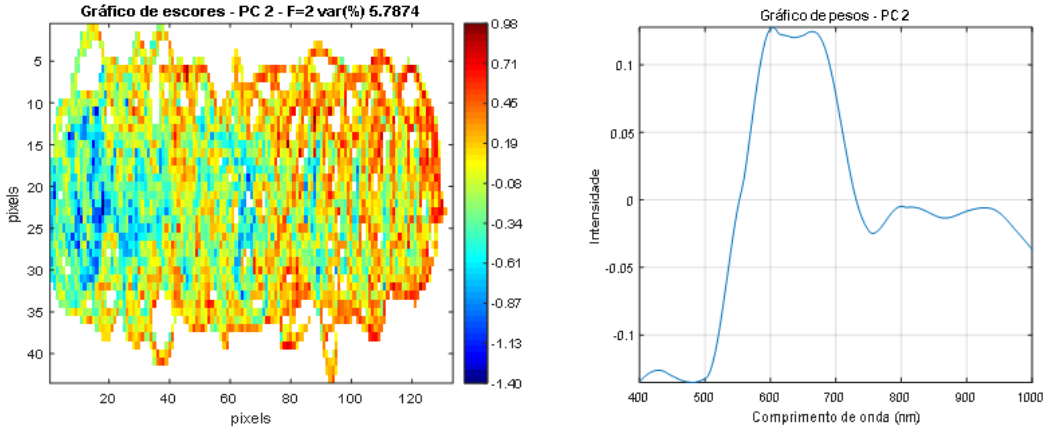
Observa-se que a figura 29, na PC 4, traz um padrão de cores que revela 4 das 5 letras da palavra "FALSO", quais sejam, as letras "F", "A", "S", "O", com uma porcentagem de 80% de acerto na identificação. Em um caso real, de aplicação forense, ainda que em um documento não sejam revelados todos os trechos obliterados por uma tinta, a identificação de trechos do texto já pode ser considerada uma informação útil para identificação de uma prova de crime. Assim, a revelação parcial da palavra, no gráfico de escores, mostra a utilidade da técnica e o gráfico de pesos demonstra que os comprimentos de onda responsáveis pela maior discriminação entre as tintas estão na faixa de 700 nm.

Outro exemplo em que foi possível verificar-se a aplicação da técnica de PCA foi amostra 4 de obliteração, na qual a palavra "FALSO" foi escrita com a tinta 1 (caneta 1) e obliterada com a tinta 2 (caneta 16).

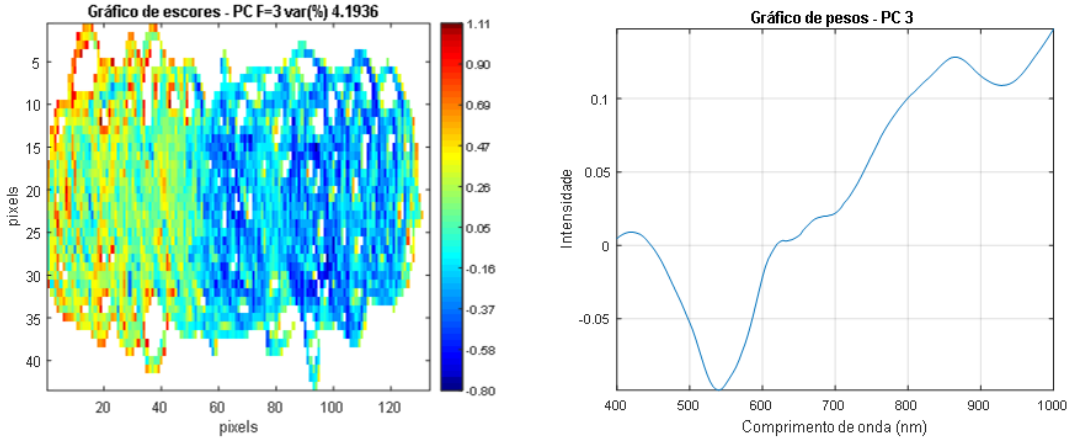
A figura 30 traz os gráficos de escores e de pesos, de (**a**) a (**d**), com os melhores resultados obtidos, com a PCA das imagens de escores e os respectivos gráficos de pesos. Com o uso de 4 componentes principais, explica-se mais de 95% da variância dos dados.



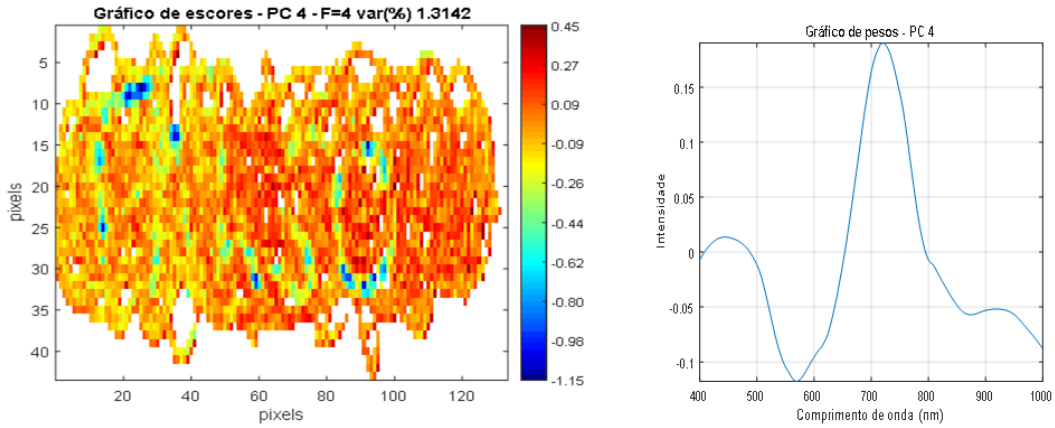
(a) Gráfico de escores (esquerda) e Gráfico de Pesos (direita) para PC 1: Percentual de variância explicada = 88%.



(b) Gráfico de escores (esquerda) e Gráfico de Pesos (direita) para PC 2. Percentual de variância explicada = 6 %.



(c) Gráfico de escores (esquerda) e Gráfico de Pesos (direita) para PC 3. Percentual de variância explicada = 4%.



(d) Gráfico de escores (esquerda) e Gráfico de Pesos (direita) para PC 4. Percentual de variância explicada = ~1 %.

Figura 30: Imagem dos escores das 4 primeiras PCS e os respectivos gráficos de pesos realizadas para a amostra 4 de obliteração: **(a) PC1; (b) PC 2;(c) PC3 e d) PC4.**

Conforme se observa na figura 30d, na PC 4, pelo padrão de escores percebido no gráfico de escores, é possível discriminar três letras da palavra "FALSO", quais sejam "F", "S" e "O", obtendo-se 60% de taxa de identificação. A análise do gráfico de pesos demonstra que a maior discriminação entre as tintas ocorre na faixa espectral de 700 nm.

Ao todo, a análise por PCA demonstrou-se satisfatória para 8 das 15 amostras de obliteração, representando 53% do total de amostras produzidas.

- **Análise Multivariada de Curvas – MCR**

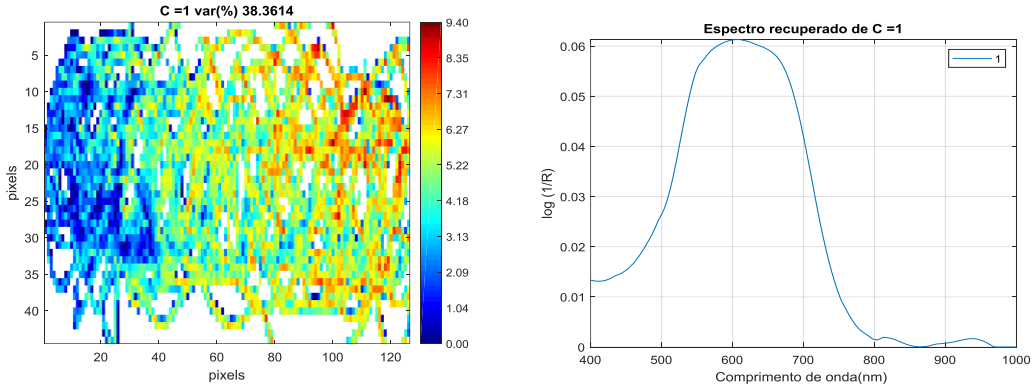
A amostra tomada como exemplo foi a amostra 6 de obliteração, na qual a palavra "FALSO" foi escrita com a tinta 1 (caneta 5) e obliterada com a tinta 2 (caneta 11).

O tratamento dos dados é semelhante ao já descrito com a técnica de análise multivariada, por PCA. São aplicados os mesmos tratamentos de redução do tamanho da imagem ("spatial cropping" e "spatial binning"), bem de pré-processamento espectral de suavização ("Smoothing Sav Gol", com janela de 15 e polinômio de grau 2)). A seleção dos pixels e dos espectros das tintas 1 e 2 é realizada pela aplicação da máscara k-means e aplica-se a técnica de PCA. Quando a técnica de PCA, pelo gráfico

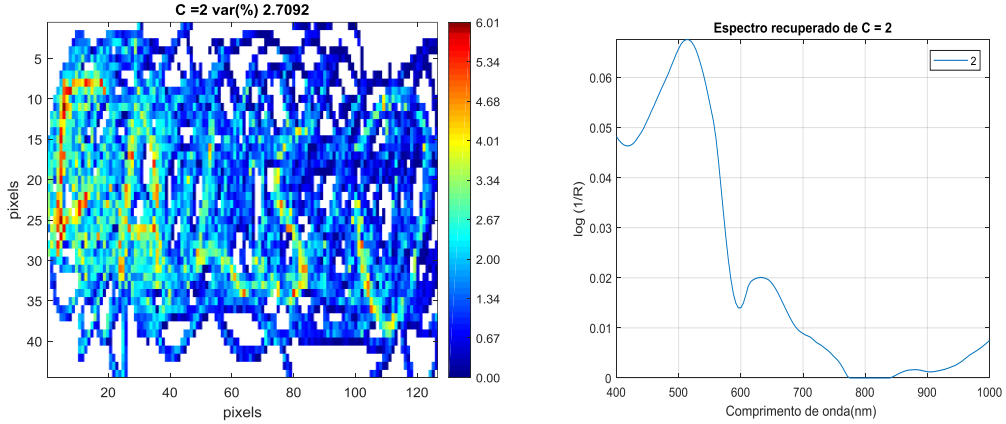
de escores e pesos, não se demonstra eficaz para revelar o texto obliterado, procede-se, então, à análise por MCR. Realiza-se a seleção das componentes pela técnica "PURITY" e aplica-se a restrição de não-negatividade.

A técnica de MCR terá discriminado a mistura de tintas se forem observadas intensidades diferentes nos valores de intensidades relativas para tintas diferentes. Se for observado, no mapa de distribuição, um padrão de cores para um trecho que revele letras da palavra FALSO, a técnica terá identificado a tinta que estava por baixo e a mistura terá sido separada.

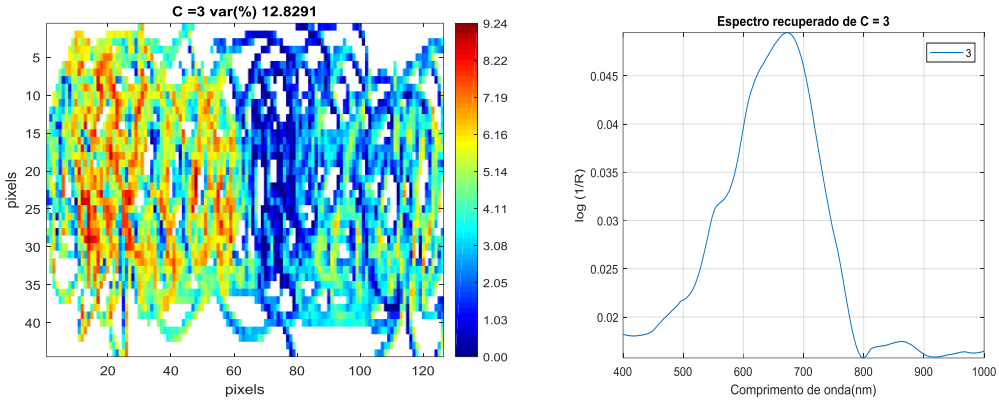
As análises por MCR foram realizados com até 6 componentes da mistura. Os resultados eram analisados e eram selecionadas as imagens das componentes que melhor discriminassem as letras. Para a amostra utilizada como exemplo, amostra 6 de obliteração, a análise foi feita com 6 componentes, as quais explicaram cerca de 70 % da variância dos dados. Os resultados podem ser observados na Figura 31, que traz os mapas de distribuição e de espectros recuperados, de (**a**) a (**f**).



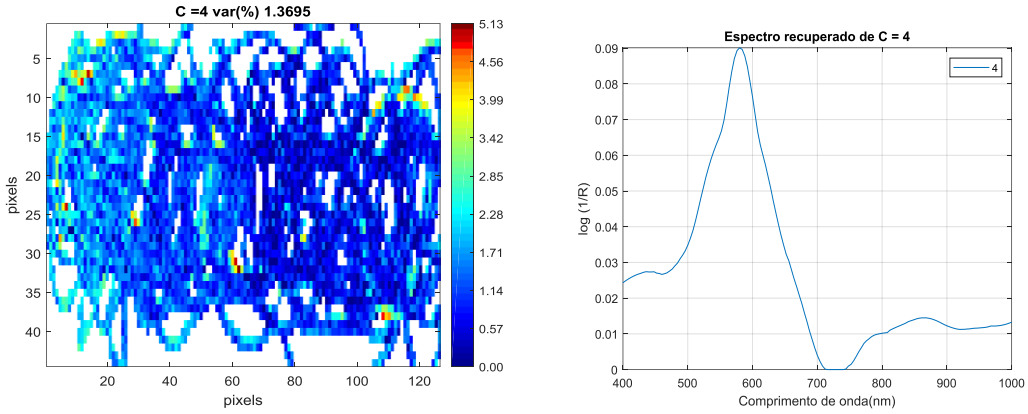
(a) Mapas de distribuição (esquerda) e Espectro puro (direita) para C=1: Percentual de variância explicada = 38%.



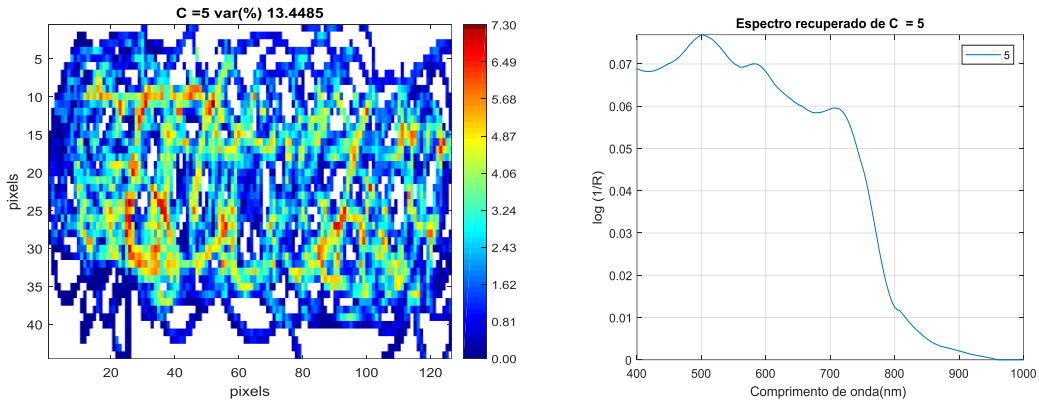
(b) Mapas de distribuição (esquerda) e Espectro puro (direita) para C= 2. Percentual de variância explicada = 3 %.



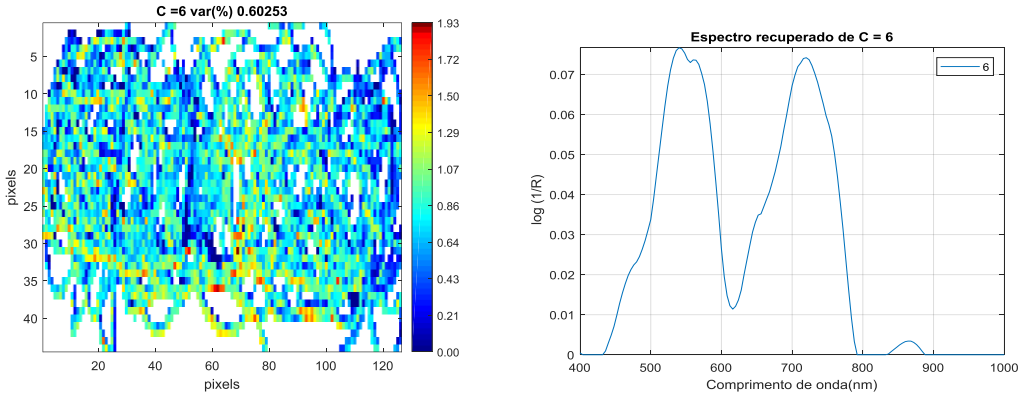
(c) Mapas de distribuição (esquerda) e Espectro puro (direita) para C= 3. Percentual de variância explicada = 13 %.



(d) Mapas de distribuição (esquerda) e Espectro puro (direita) para C= 4. Percentual de variância explicada = 1 %.



(e) Mapas de distribuição (esquerda) e Espectro puro (direita) para C= 5. Percentual de variância explicada = 13 %.



(f) Mapas de distribuição (esquerda) e Espectro puro (direita) para C= 6. Percentual de variância explicada = 1 %.

Figura 31: Imagem dos mapas de distribuição das 6 primeiras componentes e os respectivos gráficos de espectros realizadas para a amostra 6 de obliteração: **(a) C=1; (b) C=2; (c) C=3, d) C=41, e) C=5 e f) C=6.**

Assim, observa-se na figura 31b uma imagem perfazendo um total de 100% de acerto, com a identificação das letras “F”, “A”, “L”, “S” e “O”. Pelo gráfico do espectro recuperado, percebe-se que a região de faixa espectral responsável pela maior diferenciação entre as tintas das canetas 5 e 11 é a correspondente à 500 nm.

Para fins de comparação, a figura 32 traz os gráficos de escores obtidos das análises por PCA, demonstrando que a análise por PCA não foi satisfatória e que a análise por MCR era necessária.

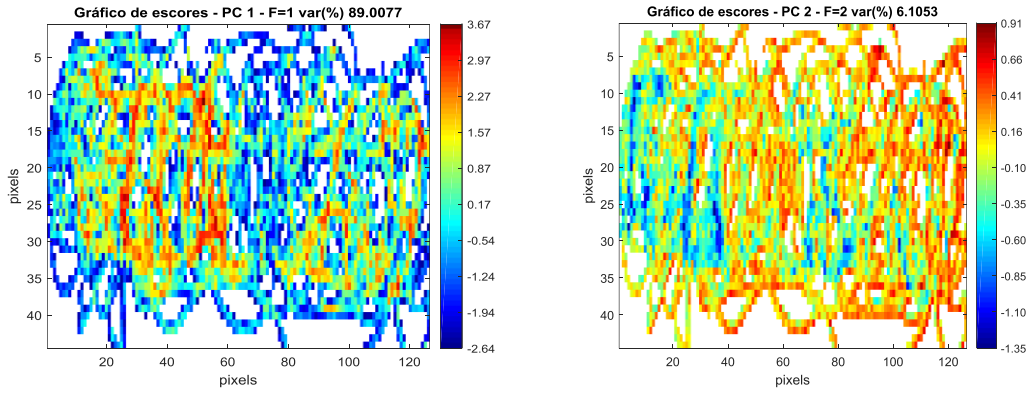
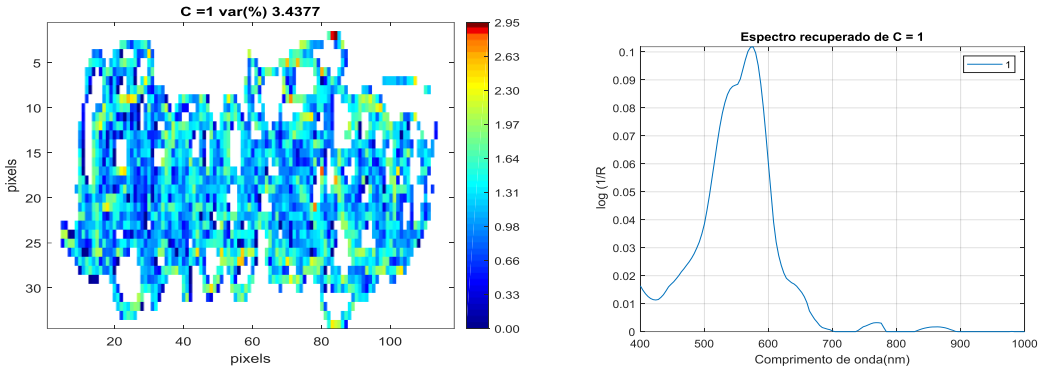
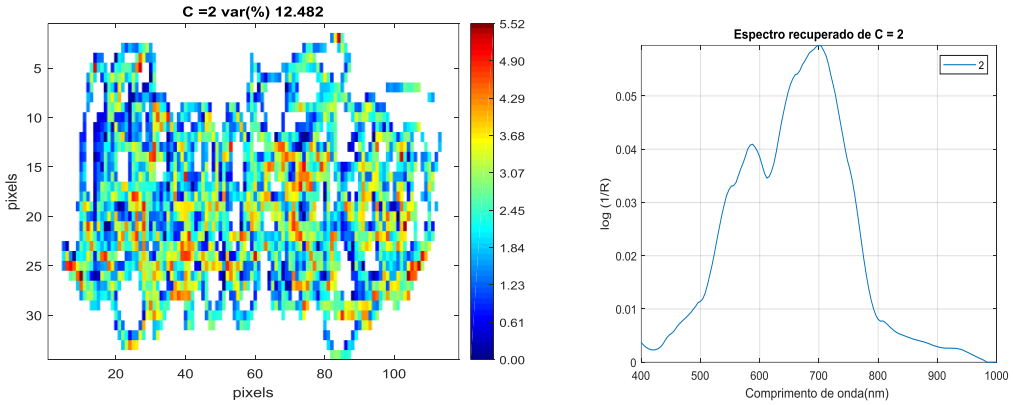


Figura 32: Gráfico de escores obtidos pelas análises de PCA, para duas componentes principais, para a amostra 6 de obliteração.

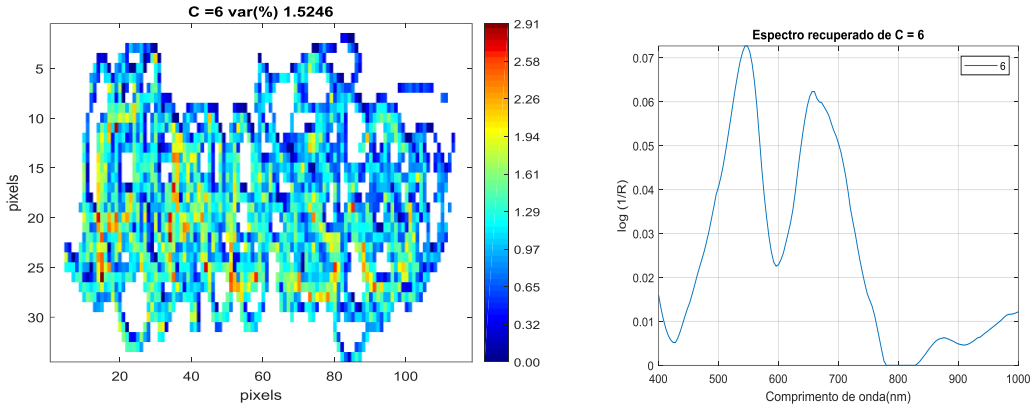
Outro exemplo em que é possível verificar-se a aplicação da técnica de MCR é amostra 13 de obliteração, na qual a palavra “FALSO” foi escrita com a tinta 1 (caneta 12) e obliterada com a tinta 2 (caneta 16). A análise dessa amostra foi feita com 6 componentes, explicando um total de 60 % da variância dos dados, e figura 33 traz os melhores resultados obtidos.



(a) Mapas de distribuição (esquerda) e Espectro puro (direita) para C=1: Percentual de variância explicada =3 %.



(b) Mapas de distribuição (esquerda) e Espectro puro (direita) para C=2. Percentual de variância explicada = 12 %.



(c) Mapas de distribuição (esquerda) e Espectro puro (direita) para C=6. Percentual de variância explicada = 2 %.

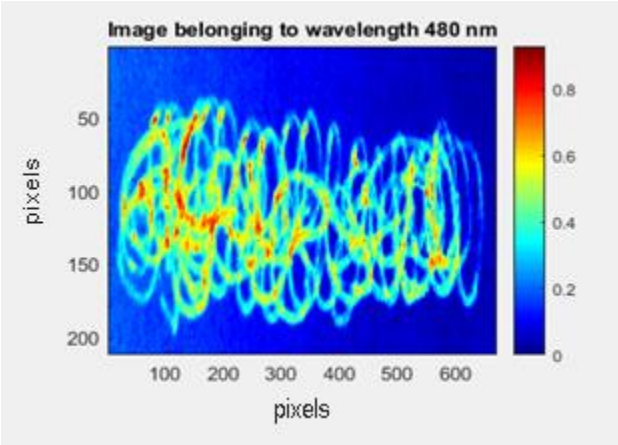
Figura 33: Imagem dos mapas de distribuição das 3 componentes selecionadas para a amostra 13 de obliteração e os respectivos gráficos de espectros: **(a) C=1; (b) C=2; (c) C=6.**

Conforme se observa na figura 33, na Componente 6, foi possível discriminar três letras da palavra "FALSO", quais sejam "F", "S" e "O", obtendo-se 60% de taxa de identificação. A análise do gráfico de pesos demonstra que a maior discriminação entre as tintas ocorre na faixa espectral entre 500 e 700 nm.

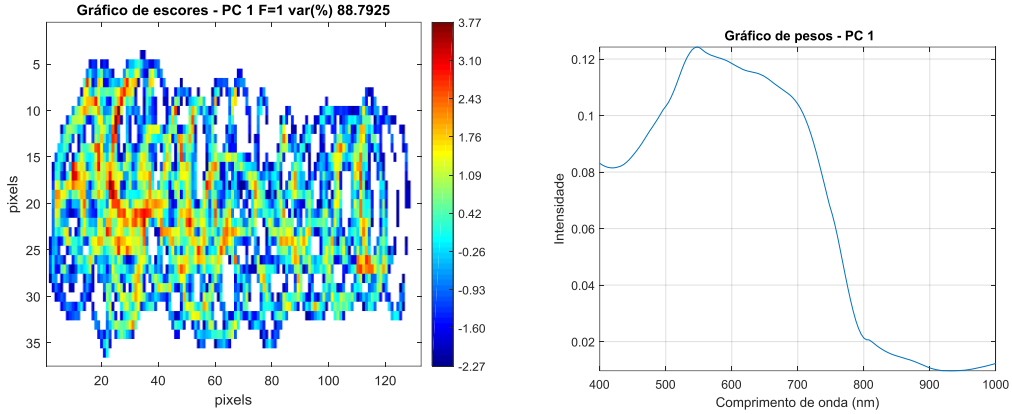
Ao todo, a análise por MCR demonstrou-se satisfatória para 3 amostras de obliteração, representando 20% do total de amostras produzidas.

- **Inconclusivo**

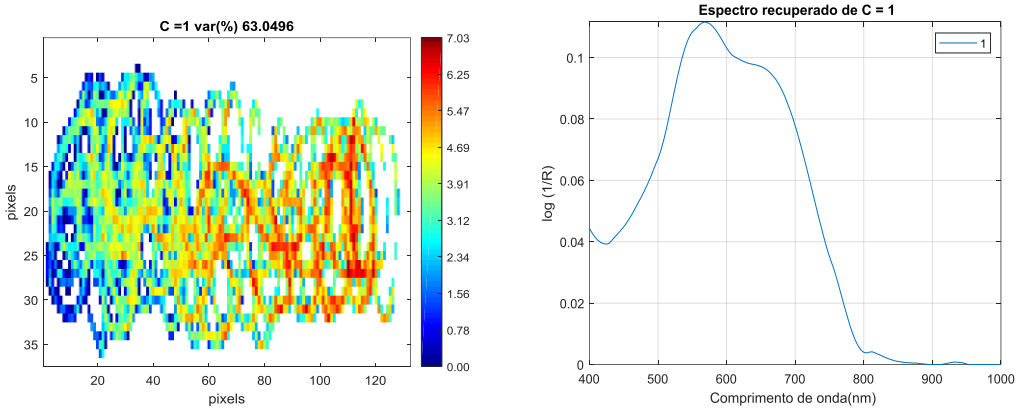
A amostra tomada como exemplo foi a amostra 11 de obliteração, na qual a palavra "FALSO" foi escrita com a tinta 1 (caneta 11) e obliterada com a tinta 2 (caneta 16). O resultado foi considerado inconclusivo quando após a aplicação o tratamento dos dados, com os pré-processamentos espaciais e espectrais já descritos nas seções anteriores e a utilização das técnicas de análise univariada e multivariada – por PCA e por MCR – não foi possível revelar nenhuma letra da palavra obliterada. A Figura 34 traz as imagens (de **a** a **c**), com os melhores resultados obtidos os quais ainda assim demonstraram-se insuficientes para a revelação de qualquer trecho das letras da palavra.



(a) Exemplo de imagem, em pixels, correspondente a comprimento de onda de 480 nm (não foi possível obter discriminação para nenhum dos comprimentos de onda).



(b) Exemplo de gráfico de escores (esquerda) e Gráfico de Pesos (direita) para PC=1: Percentual de variância explicada = 89%.



(c) Exemplo de mapas de distribuição (esquerda) e Espectro puro (direita) para C=1: Percentual de variância explicada =63%.

Figura 34: Exemplos de análises inconclusivas para amostra 11 de obliteração: **(a) Univariada (b) Multivariada - PCA (c) Multivariada – MCR.**

A impossibilidade de revelação da letra pode ter ocorrido por excesso de tinta utilizada para a produção das amostras ou pela semelhança na composição das tintas das canetas misturadas.

- **Resumo das Análises de Obliteração**

Foram analisadas 15 amostras de obliteração, para as quais o método de análise univariada foi eficaz para 20% dos casos e o de análise multivariada, para 73% dos casos, sendo 53% com a utilização de PCA e 20% com o uso de MCR. Para um total de 7% dos casos, não foi possível revelar nenhuma letra que estava escondida, gerando um resultado inconclusivo para a detecção da falsificação do documento. É possível que a impossibilidade de revelação da letra tenha ocorrido pela quantidade de tinta utilizada para a produção das amostras ou ainda pela semelhança na composição das tintas das canetas.

A tabela 8 traz uma compilação de todos os resultados obtidos para as amostras de obliteração.

Tabela 8. Total de resultados das análises - amostras de obliteração.

Nº da amostra	Canetas utilizadas	Método*	Nº letras identificadas	% acerto	Resultado
1	1 e 5	PCA	5	100	Resolvido
2	1 e 11	PCA	5	100	Resolvido
3	1 e 12	PCA	3	60	Parcial
4	1 e 16	PCA	3	60	Parcial
5	1 e 17	PCA	4	80	Resolvido
6	5 e 11	MCR	5	100	Resolvido
7	5 e 12	MCR	5	100	Resolvido
8	5 e 16	PCA	4	80	Resolvido
9	5 e 17	PCA	4	80	Resolvido
10	11 e 12	Uni	5	100	Resolvido
11	11 e 16	Inconclusivo	0	0	Inconclusivo
12	11 e 17	Uni	5	100	Resolvido
13	12 e 16	MCR	3	60	Parcial
14	12 e 17	PCA	3	60	Parcial
15	16 e 17	Uni	5	100	Resolvido
Total		**Uni (3)**		**20%**	
		PCA (8)		**53%**	
		MCR (3)		**20%**	
		Inconclusivo (1)		**7%**	

*Legenda: Uni=univariado; PCA= Análise de Componentes Principais e MCR= Resolução Multivariada de Curvas. (Resolvido) de 100 a 80% de revelação do texto, considerado como

caso resolvido satisfatoriamente; (Parcial) de 60 a 20% de revelação do texto, considerado como caso parcialmente resolvido; (Inconclusivo) nenhuma letra foi identificada.

- **Teste Cego**

O teste cego foi aplicado a pessoas leigas, que não tiveram treinamento prévio para a interpretação das imagens e que não tinham conhecimento do texto que estava escrito, apenas que a imagem poderia conter algum caractere numérico ou alfanumérico. Foram mostradas as melhores imagens obtidas pelas análises univariada por PCA, e por MCR, sendo solicitado ao entrevistado que escolhesse a imagem mais nítida e indicasse se existia algum caractere numérico ou alfanumérico na imagem. Os resultados obtidos para as 15 amostras estão compilados na Tabela 9.

Tabela 9. Total de resultados das análises - amostras de obliteração - Teste Cego.

Nº da amostra	Canetas utilizadas	Método*	Nº letras identificadas	% acerto	Resultado
1	1 e 5	PCA	5	100	∨
2	1 e 11	PCA	5	100	∨
3	1 e 12	Inconclusivo	0	0	x
4	1 e 16	PCA	2	40	~
5	1 e 17	PCA	5	100	∨
6	5 e 11	MCR	3	60	~
7	5 e 12	MCR	3	60	~
8	5 e 16	MCR	4	80	∨
9	5 e 17	Inconclusivo	0	0	x
10	11 e 12	Uni	5	100	∨
11	11 e 16	Inconclusivo	0	0	x
12	11 e 17	Uni	5	100	∨
13	12 e 16	Inconclusivo	0	0	x
14	12 e 17	Inconclusivo	0	0	x
15	16 e 17	Uni	5	100	∨
Total		**Uni (3)**		**20%**	∨ (3)
Total		**PCA (4)**		**27%**	∨ (3) ~ (1)
Total		**MCR (3)**		**20%**	∨ (1) ~ (2)
Total		**Inconclusivo (5)**		**33%**	x (5)

*Legenda: Uni=univariado; PCA= Análise de Componentes Principais e MCR= Resolução Multivariada de Curvas. (Resolvido ou ∨): de 100 a 80% de revelação do texto, considerado como caso resolvido satisfatoriamente; (Parcial ou ~): de 60 a 20% de revelação do texto,

considerado como caso parcialmente resolvido; (Inconclusivo ou x): nenhuma letra foi identificada.

O teste cego demonstrou que o conhecimento prévio do texto obliterado pode induzir a identificação das letras no resultado final. Para casos reais de análise forense, essa informação nunca está disponível, de forma que o teste cego deve representar melhor o percentual de acertos em situações reais.

O percentual de acertos no teste realizado pela pesquisadora foi de 93%, sendo 20% com abordagem univariada e 73% com abordagem multivariada. No teste cego, esse percentual foi de 67%, sendo 20% com a abordagem univariada e 47% com a abordagem multivariada. Considerando a eficácia como o percentual de acertos, a diferença na eficácia entre os resultados obtidos é de 26%. Isso demonstra que, uma vez que o texto escondido é o problema a ser resolvido pelo método de análise, o resultado mais fidedigno é o do teste cego, e que, em 67% de casos que são simulados problemas reais de fraude por obliteração, o método proposto pela pesquisa alcança o seu objetivo.

6. Conclusões

Demonstrou-se que, por meio de um método objetivo e com protocolo simples, utilizando-se uma técnica não–destrutiva, baseada em imagem hiperespectral e análise multivariada, e que emprega um equipamento disponível em muitas das unidades de perícia do pais foi possível a discriminação de tintas de canetas e a identificação de fraudes em noventa e dois por cento das amostras produzidas para simular casos de adição de texto e noventa e três por cento das amostras de obliteração, produzidos com canetas esferográficas de cor azul.

Para os casos de adição de texto, em um conjunto de dezessete canetas de marcas e modelos diferentes, o método de análise univariada permitiu a solução a discriminação das tintas analisadas em cinquenta por cento dos casos e que a análise multivariada foi capaz de ampliar a aplicação da técnica, promovendo a identificação da fraude quando a análise univariada não foi eficaz, em quarenta e dois por centos dos casos, sendo trinta e seis por cento com a utilização da PCA e seis por cento por meio da MCR. Para oito por cento dos casos, não foi possível discriminar as canetas, gerando um resultado inconclusivo para a detecção da fraude por adição de texto. Uma causa possível para a impossibilidade de discriminação é a similaridade da composição das tintas das canetas escolhidas.

Semelhantemente, para os casos de obliteração, a pesquisa demonstrou, para o conjunto de seis canetas de seis marcas e modelos diferentes, por meio de um modelo semi-quantititativo de análise, em que se obtém um resultado em função da quantidade de letras reveladas, que o método de análise univariada foi eficaz para vinte por cento dos casos e que, pela análise multivariada, houve eficácia para setenta e três por cento dos casos, sendo cinquenta e três por cento com a utilização da PCA e vinte por cento com o uso da MCR. Para um total de sete por cento dos casos, não foi possível revelar nenhuma letra que estava escondida, gerando um resultado inconclusivo para a detecção da fraude por obliteração. A impossibilidade de revelação da letra pode ter ocorrido pela quantidade de tinta utilizada para a produção das amostras ou pela semelhança na composição das tintas das canetas. Nos casos de obliteração, a realização de um teste cego demonstrou que existe influência no resultado obtido em função do conhecimento prévio da palavra obliterada. Mas, mesmo no teste cego, foi possível obter uma eficácia

de sessenta e sete por cento, e demonstrou-se que, em situações que representam melhor um caso real forense, o método proposto alcança o seu objetivo.

A análise multivariada com o uso de dados oriundos de imagens hiperespectrais tem sido reportada na literatura como uma ferramenta apropriada para a discriminação de tintas de caneta e, foi possível demonstrar, por meio da presente pesquisa que, com a utilização das ferramentas quimiométricas adequadas, é possível extrair, a partir da análise de um equipamento com baixa resolução espectral como o VSC6000®, informação útil à identificação de fraudes em documentos por adição de texto e por obliteração, potencializando o uso dessa técnica e aumentando a sua eficácia para a solução de casos forenses no Brasil.

7. Referências

1. Perruso, Carlos Renato et al. (Coord.). *Guia de serviços da perícia criminal federal: uma visão panorâmica: a verdade e a justiça pela ciência forense.* Brasília: Departamento de Polícia Federal. **2011.**

2. Romão, W. et al.*Quím. Nova*, 1717-1728,. **2011**.

3. Mendes, L. *Documentoscopia - Tratado de Perícias Criminalísticas.* Campinas: Millennium. **2015**.

4. Silva, V. A.. *J. Braz. Chem. Soc.*, 1552-1564, **2014**.

5. Borba, R. S.*Foren. Scienc. Inter*. 249, 73. **2015**

6. Pereira, et al. *Microc. Journ.* 130 , 412–419. **2017**.

7. Borba, T. J.. *Analy*. 142, 1106-1118. **2017.**

8. Calcerrada, M., & García-Ruiz, C. *Anal Chim Acta*, 853:143-166. **2015.**

9. Lednev, et al. *Anal. Chem.* 87, 306. **2015**.

10. Kumar, R., Kumar, V., & Sharma, V. *Spectrochim. Acta Pt A Mol. Biomol. Spectrosc.*170,19-28. **2017**

11. Kumar, R., Kumar, V., & Sharma, V.. *Spectrochim. Acta Pt A Mol. Biomol. Spectrosc.*, 175, 67-75. **2017**

12. José Augusto Da-Col, W. F. *Química Nova*, 345-354. **2018**

13. Edelman et al. *Forensic Sci. Int.*, 223, 28. **2012**.

14. Pimentel et al. *Analyst*, pp. 139, 5176. **2014.**

15. Relatório de Gestão do exercício de 2016 – Unidade Prestadora de Contas: Polícia Federal. **2016**. Disponível em http://www.pf.gov.br/institucional/acessoainformacao/auditorias/prestacao-de-contas/prestacao-de-contas-2016/relatorio-de-gestao-consolidado.pdf.

16. Aline Thaís Bruni, J. A. *Fundamentos de química forense:uma análise prática da química que soluciona crimes.* Millennium Editora. **2012**.

17. VSC6000®/HS User Manual (hardware). Foster + Freeman. **2012**.

18. Erick Simoes da Camera e Silva, S. F. *Documentoscopia: aspectos cientificos, técnicos e jurídicos.* Campinas, SP. : Millenium. **2013**.

19. M. J. Khan et al. *IEEE Acces.* 6. **2018**

20.N., M., S., S., R., D., & M., G. *ASME. J Biomech Eng.*, 2, 140. **2018.**

21. Amigo et al. *Anal Chim Acta* , 896, 34. **2015.**

22. Anna de Juan, J. J. *Anal. Methods*, 6, 4964. **2014.**

23. Março et al. . *Quím. Nova*, Vol. XY, No. 00, 1-8, 200_.**2014.**

24. Kowaslski, B. K.-C. *Chem. Ind.*, 22,882. **1978.**

25. Valderrama, P., Braga, J. W., & Poppi, R.. *Quim. Nova*, 32, 1278. **2009**.

26. Ferreira, M. M. *Quimiometria-Conceitos, Métodos e Aplicações.* Campinas, SP: Editora da Unicamp. **2015.**

27. N. Mobaraki, J. A. *Chemom. Intell. Lab. Syst*,172 ,174–187. **2018.**

28. J. A. *Chemom. Intell. Lab. Syst* ,117 138 –148 145M. **2012**.

29. Souza, A., & Poppi, R. *Quim. Nova*, 35, 223. **2012.**

30. José Augusto Da-Col, W. F. *Quim. Nova*, 3, 345-354. **2018.**

31. Brereton, R. *Chemometrics for Pattern Recognition.* Chichester: John Wiley & Sons. **2007.**

32. Wold, S., Esbensen, K., & Geladi, P. *Chemom. Intell. Lab. Syst.*, 2, 37. **1987.**