Uma Técnica para Alinhamento de Imagens de Documentos Antigos

Valguima V. V. A. Odakura Martinez[1] **, Geraldo Lino de Campos**[1]

[1]Departamento de Engenharia de Computação e Sistemas Digitais
Escola Politécnica da Universidade de São Paulo
Av. Prof. Luciano Gualberto, 158, tv. 3 – 05508-900 São Paulo, SP

valguima.odakura@poli.usp.br, geraldo.campos@poli.usp.br

Abstract. *This work describes a technique to ancient document images alignment. To be possible to align two images of the same document page, an image registration approach is used. These images have acquisition differences as translation, rotation and distortions. The first step of registration is the control points selection. In this stage, images of the page are segmented in text lines and then in words. Therefore, each selected control point corresponds to the begin of a word. In the second stage, a matching is performed between extracted control points sets of both images because points selection presents faults and both sets are not equals. In the last stage, target image is mapped using thin plate spline functions to coincide it with the reference image.*

Resumo. *Este trabalho descreve uma técnica para alinhar imagens de documentos antigos. Para conseguir alinhar duas imagens de uma mesma página uma técnica de registro de imagens é utilizada. As imagens possuem diferenças de aquisição como translação, rotação e distorções. A fase inicial para o registro é a seleção de pontos de controle. Nesta fase, as imagens da página são segmentadas em linhas de texto e depois em palavras, sendo que cada ponto de controle equivale ao início de uma palavra. Na segunda fase os dois conjuntos de pontos de controle são correspondidos, pois a seleção de pontos apresenta falhas e os conjuntos são diferentes. Na fase final, a imagem destino é mapeada usando funções thin plate spline para que coincida com a imagem de referência.*

1. Introdução

Um dos principais problemas da digitalização de livros antigos é o tipo de papel utilizado naquela época, que era mais transparente que aquele utilizado hoje e, por isso, deixava visível a impressão do verso da folha. Tal problema pode não parecer grave para leitores do documento original, uma vez que o olho humano consegue distinguir entre as marcas da folha da frente e as marcas do verso da folha. Entretanto, quando a imagem é digitalizada e exibida em um monitor, as marcas do verso da folha aparecem como ruído e dificultam bastante a leitura, tornando necessário realizar um tratamento da imagem para eliminar tais marcas do verso.

Para remover o efeito da transparência, segundo [Stolfi, 2000], cada página do livro deve ser digitalizada duas vezes, uma sobre um fundo branco e outra sobre um fundo

preto. Ao digitalizar duas vezes a mesma página do documento as imagens resultantes podem apresentar diferenças. Intuitivamente, ao digitalizar uma página do livro sobre uma folha preta, o efeito da transparência é praticamente eliminado, e ao digitalizar a mesma página sobre uma folha branca, o efeito da transparência é realçado. Essas duas imagens possuem diferenças e devem ser alinhadas, para em seguida, eliminar o efeito da transparência.

Ao digitalizar duas vezes a mesma página do documento, as imagens resultantes podem apresentar diferenças em relação à inclinação e à posição no plano. Esse processo faz com que as coordenadas dos pontos em uma imagem não coincidam com as coordenadas na outra imagem, ou seja, o mesmo par de coordenadas nas duas imagens não represente a mesma porção do documento. Deseja-se então alinhar essas imagens de forma que suas coordenadas sejam correspondentes. Existem ainda diferenças devido ao processo de manuseio das folhas do livro pode produzir deformações diferentes em partes distintas da folha.

Para que o alinhamento seja possível há ainda uma outra diferença presente nas imagens que precisa ser tratada. Como as imagens resultam de livros, que em geral são volumosos, as imagens podem apresentar distorções. Dado um livro volumoso aberto, a distorção pode aparecer no meio do livro, mais precisamente em uma porção na junção da página do lado esquerdo com a página do lado direito. A distorção pode ainda ocorrer no extremo esquerdo da página esquerda e no extremo direito da página direita. Desta forma, o objetivo deste trabalho é realizar o alinhamento de duas imagens da mesma página de um documento, que possuem diferenças devido a sua aquisição.

Na Seção 2. o problema do registro de imagens é introduzido, como uma técnica para realizar o alinhamento das imagens. A Seção 3. descreve a técnica de registro de imagens utilizando funções *thin plate spline*. A seleção de pontos de controle, baseada em técnicas de processamento de documentos, é apresentada na Seção 4.. A correspondência dos conjuntos de pontos de controle é tratada na Seção 5.. Os resultados obtidos com a técnica descrita são mostrados na Seção 6.. Por fim, na Seção 7. a conclusão do trabalho é apresentada.

2. Registro de Imagens

O registro de imagens é uma tarefa comum em aplicações de processamento de imagem e visão computacional. O registro é utilizado para coincidir duas ou mais imagens obtidas de diferentes sensores, em diferentes tempos ou de diferentes pontos focais. Um grande número de métodos de registro de imagens pode ser encontrado na literatura, como por exemplo em [Brown, 1992] e [den Elsen et al., 1993]. Qualquer dos métodos produz um conjunto de equações que transformam as coordenadas de cada ponto de uma imagem em coordenadas do ponto correspondente na outra imagem. A tarefa do registro é, então, determinar como transformar a primeira imagem de forma que ela coincida com a segunda.

São consideradas duas imagens da mesma cena que apresentam diferenças devido ao processo de aquisição, representando a imagem de referência e a imagem destino. O **registro de imagens** pode ser definido como sendo um mapeamento entre as duas imagens de forma que essas diferenças sejam minimizadas. As funções de transformação entre as imagens são f_x e f_y como segue:

$$u = f_x(x, y) \quad \text{e} \tag{1}$$
$$v = f_y(x, y). \tag{2}$$

Tais funções relacionam as coordenadas da imagem de referência na forma (x, y) com as coordenadas da imagem destino na forma (u, v).

O processo de registro de imagens pode ser dividido em três passos: seleção dos pontos de controle, correspondência desses pontos e estimativa da função de mapeamento. No primeiro deles, são determinados dois conjuntos de pontos de controle. Ou seja, são determinados dois subconjuntos de pontos, um da imagem I_1 e outro da imagem I_2, denotados por C_1 e C_2. No segundo, os pontos em C_1 devem ser correspondentes aos pontos em C_2, isto é, para cada ponto em C_1 que representa uma porção da imagem existe um ponto em C_2 que representa a mesma porção da imagem. Esses pontos são chamados de **pontos de controle**. No terceiro passo, esses pontos correspondentes são utilizados para estimar uma função de mapeamento que possa relacionar os pontos restantes nas imagens. A função de mapeamento é escolhida entre os vários tipos de transformações.

3. Registro de Imagens Utilizando Funções *Thin Plate Spline*

Devido ao fato de algumas imagens apresentarem distorções complexas como, por exemplo, projeção bidimensional de objetos tridimensionais, movimentos de objetos incluindo efeitos de oclusão e as deformações de objetos elásticos, existe a necessidade de encontrar métodos capazes de registrar imagens com tais distorções. Os métodos que solucionam esse problema exploram modelos elásticos. Neste trabalho, como o processo de manuseio das folhas de um livro pode produzir deformações diferentes em diferentes partes da folha, decidiu-se pela utilização da transformação elástica, que permite corrigir esse tipo de deformação, além da translação e rotação que ocorrem ao serem realizadas duas digitalizações com fundos diferentes.

Um modelo elástico para o alinhamento de imagens é o *Thin Plate Spline – TPS* [Goshtasby, 1988]. Os resultados mais precisos em registro de imagens com distorções geométricas locais foram obtidos usando as funções de mapeamento de superfície *spline*, segundo [Barrodale et al., 1992]. Por essa razão, o método de registro de imagens utilizado nesse trabalho será o TPS.

O problema de encontrar as funções de mapeamento f_x e f_y utilizando TPS pode ser formulado como um problema de interpolação de superfícies de maneira que as superfícies obtidas representam as compontes f_x e f_y da função de mapeamento e caracterizam as distorções geométricas entre as imagens.

Desta forma, dados dois conjuntos de pontos tridimensionais

$$S = \{(x_i, y_i, u_i): \quad \text{para } i = 1, \ldots, n\} \quad \text{e} \tag{3}$$
$$Q = \{(x_i, y_i, v_i): \quad \text{para } i = 1, \ldots, n\}, \tag{4}$$

quer-se encontrar superfícies suaves $f_x(x, y)$ e $f_y(x, y)$. A superfície $f_x(x, y)$ deve passar por todos os pontos em S e a superfície $f_y(x, y)$ deve passar por todos os pontos em Q.

As superfícies desejadas podem ser encontradas através das expressões:

$$f_x(x,y) = a_0 + a_1 x + a_2 y + \sum_{i=1}^{n} F_i r_i^2 \log r_i^2 \quad \text{e} \tag{5}$$

$$f_y(x,y) = b_0 + b_1 x + b_2 y + \sum_{i=1}^{n} G_i r_i^2 \log r_i^2, \tag{6}$$

onde $r_i^2 = (x - x_i)^2 + (y - y_i)^2$.

Os coeficientes a_0, a_1, a_2 e F_i, para $i = 1, \ldots, n$, são determinados pela substituição dos n pontos de controle na Equação (5) e solucionando o sistema de $(n+3)$ equações lineares a seguir

$$f(x_i, y_i) = u_i, \quad \text{para } i = 1, \ldots, n, \tag{7}$$

$$\sum_{i=1}^{n} F_i = 0, \quad \sum_{i=1}^{n} F_i x_i = 0, \quad \text{e} \quad \sum_{i=1}^{n} F_i y_i = 0. \tag{8}$$

A superfície $f_x(x,y)$ obtida representa a primeira componente da função de mapeamento. A superfície $f_y(x,y)$, que representa a segunda componente, é determinada de maneira similar. Depois que os coeficientes foram determinados, as duas funções de mapeamento f_x e f_y desejadas foram encontradas, é possível registrar as imagens. Para efetuar a transformação aplicam-se as funções f_x e f_y a cada ponto da imagem destino.

4. Seleção de Pontos de Controle

O passo inicial para a solução do problema de registro de imagens é a seleção dos pontos de controle. Para o problema do registro de imagens de documentos foi estudada uma forma de seleção de pontos de controle baseada em técnicas de processamento de documentos. As imagens utilizadas para testes foram obtidas do livro *Sermão no Auto da Fé em Coimbra* [Mendonça, 1619] e foram digitalizadas utilizando um aparelho de *scanner* de mesa, com resolução 300×300 pixels por polegada, intensidade de 8 bits por pixel, com 256 níveis de cinza e dimensões 507×736 pixels para imagem de uma página.

A seleção de pontos de controle foi realizada supondo que as páginas do documento possuem somente linhas de texto, sem figuras. Se a página contiver figuras, a seleção de pontos de controle empregada não funciona. Para o caso de páginas com figuras, deve-se antes segmentar a imagem de forma que as partes textual e não textual sejam separadas. Então pode-se aplicar a técnica descrita a seguir para a parte que contém somente texto.

Como as imagens utilizadas neste trabalho para testes possuem 256 níveis de cinza, deve-se calcular o histograma de cada imagem e encontrar um limiar para cada uma delas. Os pixels da imagem que tiverem um valor menor ou igual ao limiar correspondente serão chamados de pixels em preto e os demais de pixels do fundo da página.

Os pontos de controle são selecionados da segmentação da imagem em linhas de texto e em seguida em palavras. A segmentação da imagem do texto será executada utilizando uma técnica semelhante àquela empregada por [Muge et al., 2000].

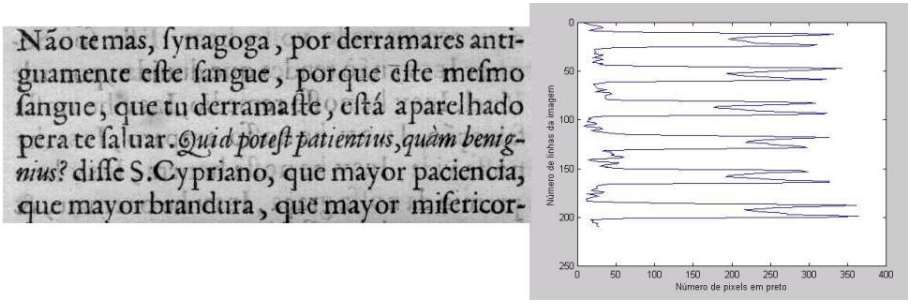


Figure 1: Porção de uma página de texto e seu respectivo histograma.

4.1. Segmentação da Página de Texto em Linhas

A segmentação da imagem em linhas de texto começa com a computação de um vetor de número de pixels em preto por linha. Com base nesse vetor, uma operação é efetuada envolvendo a computação de um limiar para extrair as linhas de texto. O vetor de número de pixels em preto é uma operação simples que percorre as linhas da imagem e para cada linha armazena o número de pixels em preto que aparecem na linha. Assim, no final dessa operação, um vetor dos valores obtidos constitui um histograma. Na Figura 1 é apresentada uma imagem com linhas de texto e a representação do seu histograma.

Ao observar a Figura 1 é possível intuitivamente identificar as linhas de texto. Percorrendo o histograma verticalmente encontram-se alternadamente picos e vales. Os picos representam as linhas de texto e os vales os espaços entre linhas. Para computar a localização de uma linha de texto, uma medida estatística chamada valor médio m de pixels por linha é introduzida. Sua computação é realizada utilizando a seguinte expressão:

$$m = \sum_{i=1}^{lin} histo_i/lin, \tag{9}$$

onde lin é o tamanho do vetor de número de pixels em preto por linha e $histo_i$, com $i = 1, \ldots, lin$, é cada elemento desse vetor.

Sendo conhecido o limiar m, percorre-se o vetor de número de pixels por linha para identificação do início e fim da linha de texto. Esta técnica deixa algumas linhas sem identificação, como é o caso de linhas de texto de comprimento menor que a média. Uma maneira de solucionar esse problema é reaplicar o mesmo algoritmo para as porções da imagem que não contribuíram com nenhuma linha.

Na presença de inclinação na imagem da página do documento, as linhas de texto não mantêm sua aparência distintiva no vetor de número de pixels em preto por linha. Uma imagem de uma página de texto inclinado e seu histograma podem ser vistos na Figura 2. Observando o histograma é fácil ver que a transição entre picos e vales não identifica o início e o fim de cada linha de texto.

Como a segmentação da página em linhas não funciona para imagens de páginas inclinadas, foi verificada a necessidade de corrigir as possíveis diferenças de inclinação antes de realizar o processo de segmentação em linhas. Existem na literatura muitos métodos para corrigir a inclinação da página, como em [Tang et al., 1996] e [Le et al., 1994].

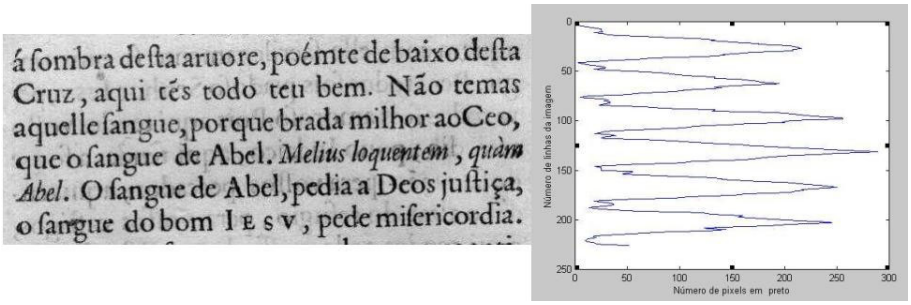


Figure 2: Porção de uma página de texto inclinada e seu respectivo histograma.

Uma vez que a inclinação da imagem da página foi corrigida, a imagem está pronta para ser segmentada em linhas, aplicando a técnica descrita anteriormente. Depois da imagem já estar devidamente segmentada em linhas, falta a segmentação de cada uma das linhas em palavras, para se obter os conjuntos de pontos de controle.

4.2. Segmentacão de Linhas de Texto em Palavras

A segmentação de um texto em palavras baseia-se em uma suposição do espaçamento do texto. Por observação de documentos, assume-se que o espaço que separa dois caracteres é menor que o espaço que separa duas palavras. Em imagens de documentos esse espaço é medido em pixels. Para detecção de palavras, assume-se também que as linhas de texto da imagem já foram identificadas, sendo a técnica de segmentação de palavras aplicada para cada uma das linhas obtidas.

Assim, o primeiro passo para a segmentação é o cálculo do vetor de número de pixels em preto por coluna de cada imagem, lembrando que agora uma imagem é uma linha de texto. O vetor de número de pixels em preto é quase igual ao anterior, ou seja, para cada coluna da imagem computa-se o número de pixels em preto presentes. A partir do vetor de número de pixels em preto por coluna é posível construir o histograma de comprimentos de espaços em branco que estão presentes em cada linha. Através desse histograma escolhe-se um limiar apropriado para divisão em palavras.

O algoritmo para segmentação em palavras pode apresentar algumas falhas se a suposição inicial de espaçamento não é verificada. Como os documentos aqui tratados são antigos, podem possuir irregularidades no formato que podem levar a falhas do algoritmo. No entanto, como a técnica de segmentação é aplicada neste trabalho com o intuito de extrair pontos de controle e não necessariamente identificar todas as palavras do texto, a perda de algumas palavras não será prejudicial para seu desenvolvimento. Devido as falhas, os conjuntos de pontos extraídos das duas imagens são diferentes, ou seja, possuem tamanhos diferentes. Desta forma, torna-se necessário realizar a tarefa de correspondência dos conjuntos de pontos encontrados.

5. Correspondência dos Pontos de Controle

Vários trabalhos, encontrados na literatura, trataram do problema da correspondência dos conjuntos de pontos controle. Neste trabalho, procura-se uma forma de corresponder os pontos de controle que possua menor custo computacional possível. Seja n o tamanho do

maior conjunto de pontos de controle. As melhores técnicas encontradas têm complexidade $O(n^2 \log n)$.

Como a extração de pontos é feita linha por linha, pode-se fazer a correspondência dos pontos também desta forma. A aplicação de uma técnica de correspondência para os pontos de cada linha, ao invés de para o conjunto inteiro de pontos diminui consideravelmente o tempo de computação. Por exemplo, suponha uma página com 25 linhas, cada linha com aproximadamente 7 palavras. Neste caso, existem 175 palavras por página. Se for utilizado o algoritmo de complexidade $O(n^2 \log n)$ para o conjunto de pontos de tamanho 175, gasta-se $175^2 \times \log 175 = 6.9 \times 10^4$ computações. No entanto, se a mesma técnica for aplicada para cada linha, têm-se $7^2 \times \log 7 \times 25 = 4 \times 10$ computações, que representa uma economia da ordem de 10^3 computações.

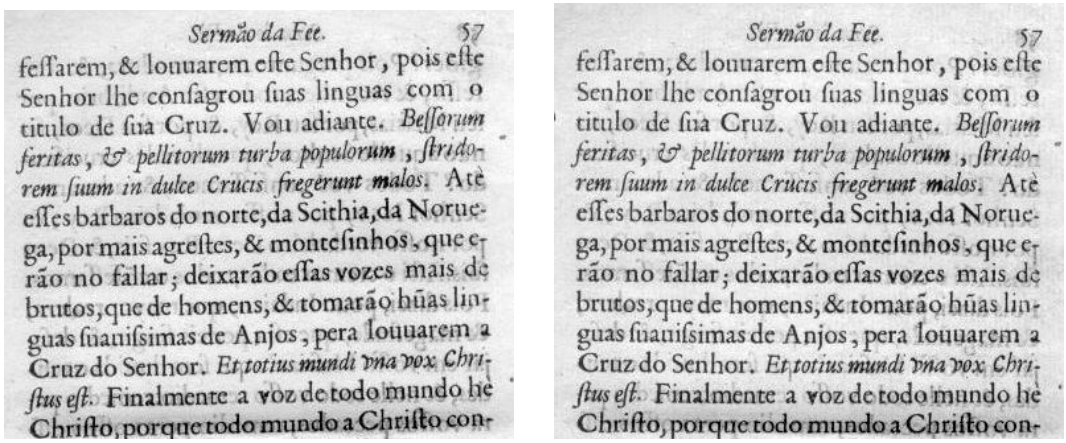
A idéia da correspondência linha por linha pode ser aplicada, neste trabalho, para qualquer técnica de correspondência existente, reduzindo consideravelmente a complexidade computacional desta tarefa. No entanto, observando a forma de seleção dos pontos de controle, feita linha por linha, e aproveitando a idéia de correpondência da mesma forma, uma nova maneira de efetuar a correspondência pode ser empregada. Levando em consideração que o conjunto de pontos encontrados para cada linha está ordenado pela coordenada y, isto é, pela coluna da imagem em que se inicia uma palavra, a correspondência dos pontos de controle pode ser feita em tempo linear.

Um método para correspondência de pontos de controle utilizando as características descritas acima será proposto a seguir. A idéia do algoritmo é, dados dois conjuntos de pontos de controle, um de cada linha, percorrê-los comparando pares de pontos. Caso a distância entre um par de pontos esteja dentro de um limiar, tais pontos são correspondidos e então passa-se para os pontos seguintes. Caso contrário, verifica-se qual dos conjuntos possui a coordenada y de menor valor, o próximo ponto desse conjunto é tomado e então repete-se o procedimento. O limiar utilizado é escolhido de acordo com a tolerância permitida para cada par correspondente.

6. Resultados

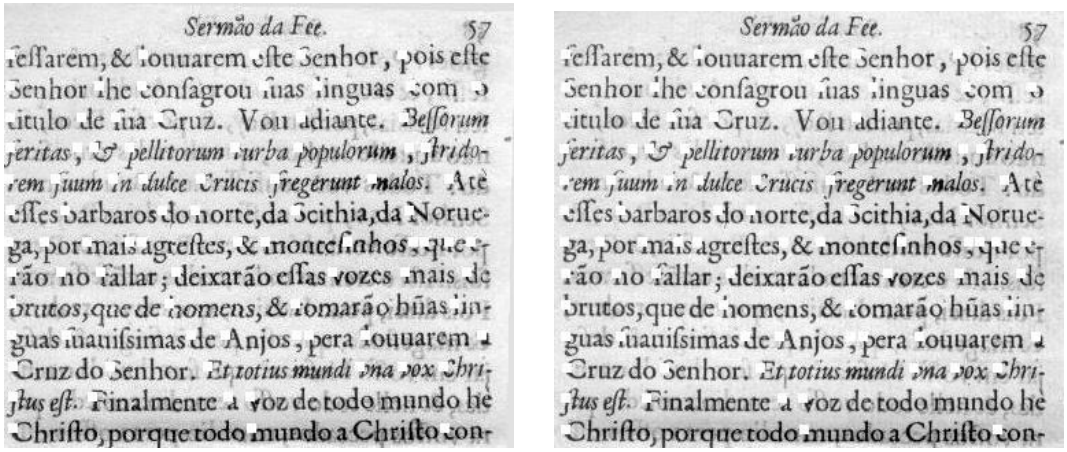
Para exemplificar o registro de imagens de documentos utilizando TPS, foi realizado um teste. Na Figura 3 podem ser vistas no item (a) a imagem de referência e no item (b) a imagem destino. A soma das duas imagens de entrada é apresentada no item (a) da Figura 6, pela qual é possível observar as diferenças entre elas. As duas imagens de entrada foram pré-processadas para que uma possível inclinação da página fosse corrigida. Na Figura 4 item (a) estão os pontos de controle para a imagem de referência e no item (b) estão os pontos de controle para a imagem destino, ambos selecionados como pontos dos inícios das palavras, marcados com pontos brancos um pouco exagerados. O registro das imagens pode ser observado na Figura 5 e a soma da imagem de referência com a imagem registrada pode ser vista no item (b) da Figura 6.

A comparação entre os itens (a) e (b) da Figura 6 mostra que as diferenças entre as imagens foram significantemente reduzidas e é considerado um bom resultado. Vale ressaltar que as deformações das imagens utilizadas foram intencionalmente muito maiores que aquelas que ocorrem em situações gerais, a fim de estressar o processo aos seus limites.



(a) (b)

Figure 3: (a) Imagem de referência digitalizada com fundo branco. (b) Imagem destino digitalizada com fundo preto.



(a) (b)

Figure 4: Pontos de controle: (a) da imagem de referência. (b) da imagem destino.

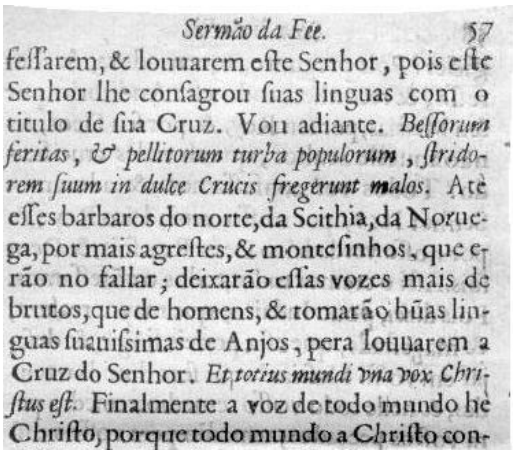
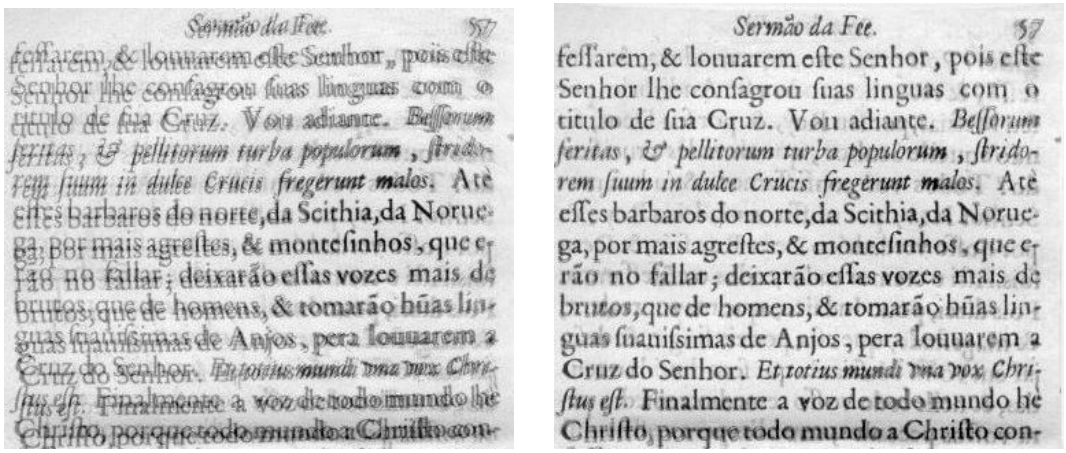


Figure 5: Imagem destino registrada.



(a) (b)

Figure 6: (a) Soma das imagens de entrada. (b) Soma da imagem registrada com imagem de referência.

Uma maneira de medir o resultado do registro de imagens encontrado é através da soma dos quadrados das diferenças entre as imagens. Essa técnica foi utilizada por [Takeuchi and Herbert, 1998] para comparar as imagens após o registro. Sejam I_1 a imagem de referência, I_2 a imagem destino e I_r a imagem destino registrada. Pode-se calcular a soma dos quadrados das diferenças, $E^2_{1,2}$, para as imagens I_1 e I_2, que representa as diferenças entre as imagens antes do registro, como segue:

$$E^2_{1,2} = \sum_{x=1,y=1}^{lin,col} [I_1(x,y) - I_2(x,y)]^2, \tag{10}$$

onde lin é o número de linhas da imagem e col é o número de colunas. Calcula-se novamente $E^2_{1,r}$, desta vez para as imagens I_1 e I_r, que representa as diferenças entre as imagens após o registro, da mesma forma.

A soma dos quadrados das diferenças para o teste apresentado resultou os valores $E^2_{1,2} = 5.5650 \times 10^6$ e $E^2_{1,r} = 1.5063 \times 10^6$. Para comparar os valores obtidos acima foi computada a razão entre $E^2_{1,r}$ e $E^2_{1,2}$, que pertence ao intervalo $[0, 1] \in \Re$ e quanto mais a razão se aproxima de 0 mais próximas são as imagens. Ao dividir $E^2_{1,r}$ por $E^2_{1,2}$ obteve-se 0.2719. Com base nesse resultado considera-se que o registro apresentou uma melhora significativa na diferença entre a imagem de referência e a imagem registrada, tornando o resultado do registro satisfatório.

7. Conclusão

A preocupação com a preservação dos livros antigos, de importância cultural ou histórica, e a divulgação do conhecimento neles encerrado motivou muitas pesquisas que conduzem à construção de bibliotecas digitais acessíveis através de redes de computadores. Tal tarefa contribui indubitavelmente para que obras raras possam ser melhor difundidas. Neste contexto, este trabalho contribui com uma pequena parte dessa empreitada, realizando o alinhamento das imagens.

Ao analisar as imagens digitalizadas obtidas, diferenças de rotação, translação e distorções foram encontradas. Para minimizar essas diferenças, foram escolhidas funções TPS como método de registro. A seleção de pontos de controle foi um problema crucial, uma vez que sem uma escolha apropriada dos pontos não é possível obter um registro satisfatório das imagens. Para encontrar pontos de controle foram estudadas técnicas de processamento de documentos. Vale ressaltar que foram encontrados na literatura poucos trabalhos relacionados.

A utilização da segmentação da imagem do documento em palavras e a extração de um ponto de controle de cada palavra foi uma escolha adequada como solução para o problema. Sabendo que a seleção dos pontos de controle é feita linha por linha, foi proposto um método para correspondência dos pontos de controle que aproveita essa informação. Esse método tem desempenho bastante satisfatório, linear no número de pontos.

References

Barrodale, I., Skea, D., Berkley, M., Kuwahara, R., and Poeckert, R. (1992). Warping digital images using thin plate splines. *Pattern Recognition*, 26(2):375–376.

Brown, L. G. (1992). A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376.

den Elsen, P. A. V., Pol, E. D., and Viergever, M. A. (1993). Medical image matching – a review with classification. *IEEE Engineering in Medicine and Biology*, 12:26–39.

Goshtasby, A. (1988). Registration of images with geometric distortions. *IEEE Transactions on Geoscience and Remote Sensing*, 25(1):60–64.

Le, D. S., Thoma, G. R., and Weichsler, H. (1994). Automated page orientation and skew angle detection for binary document images. *Pattern Recognition*, 27(10):1325–1344.

Mendonça, F. (1619). *Sermão no Auto da Fé em Coimbra*. Na oficina de Diogo Gomez de Loureyro.

Muge, F., Granado, I., Mengucci, M., Pina, P., Ramos, V., Sirakov, N., Pinto, J. R. C., Marcolino, A., Ramalho, M., Vieira, P., and Amaral, A. M. (2000). Automatic feature extraction and recognition for digital access of books of the renaissance. In Borbinha, J. and Baker, T., editors, *Research and Advanced Technology for Digital Libraries*, volume 1923 of *Lecture Notes in Computer Science*, pages 24–34. Springer-Verlag.

Stolfi, J. (2000). Removing optical bleedthrough from imaged documents. A ser publicado.

Takeuchi, Y. and Herbert, M. (1998). Finding images of landmarks in video sequences. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'98)*.

Tang, Y. Y., Lee, S., and Suen, C. Y. (1996). Automatic document processing: a survey. *Pattern Recognition*, 29(12):1931–1952.