

Piotr FABIAN, Jacek OLEARCZYK  
Politechnika Śląska, Instytut Informatyki

## AUTOMATYCZNA KOREKCJA ZDJĘĆ DOKUMENTÓW

**Streszczenie.** Tradycyjnym sposobem przenoszenia drukowanych dokumentów do pamięci komputera jest skanowanie. Szybszym sposobem jest użycie cyfrowego aparatu fotograficznego. Wtedy jednak obraz może zostać zniekształcony, a oświetlenie może być nierównomierne, co utrudnia lub uniemożliwia dalsze przetwarzanie. Zaproponowana metoda automatycznej korekcji geometrii i oświetlenia pozwala poprawić jakość zdjęcia.

**Słowa kluczowe:** korekcja geometrii, korekcja oświetlenia, skanowanie

## AUTOMATIC CORRECTION OF DOCUMENT IMAGES

**Summary.** The traditional way of transferring printed documents into the computer memory is scanning. A quicker way is to use a digital camera. But in this case the picture geometry may be distorted, and lighting may be uneven. This makes further processing difficult or impossible. The proposed method for automatic correction of geometry and lighting can improve the picture quality.

**Keywords:** geometry correction, lighting correction, scanning

### 1. Wprowadzenie

Tworzenie cyfrowych kopii dokumentów dostępnych w postaci papierowej ma wiele uzasadnień, wśród nich wygodniejszy dostęp, mniejszą objętość nośnika, możliwość łatwej archiwizacji w wielu kopiach, możliwość wyszukiwania treści po odpowiednim przetworzeniu. Standardową metodą zamiany drukowanych dokumentów na postać elektroniczną jest skanowanie. Przy odpowiednim doborze sprzętu i parametrów skanowania jakość uzyskanej elektronicznej kopii jest porównywalna z oryginałem. Uzyskanie tak dobrej jakości jest jednak czasochłonne. Wymaga też specjalnego przygotowania materiałów źródłowych. Znacznie szybszą metodą, nawet 30-krotnie, jest zastosowanie cyfrowego aparatu fotogra-

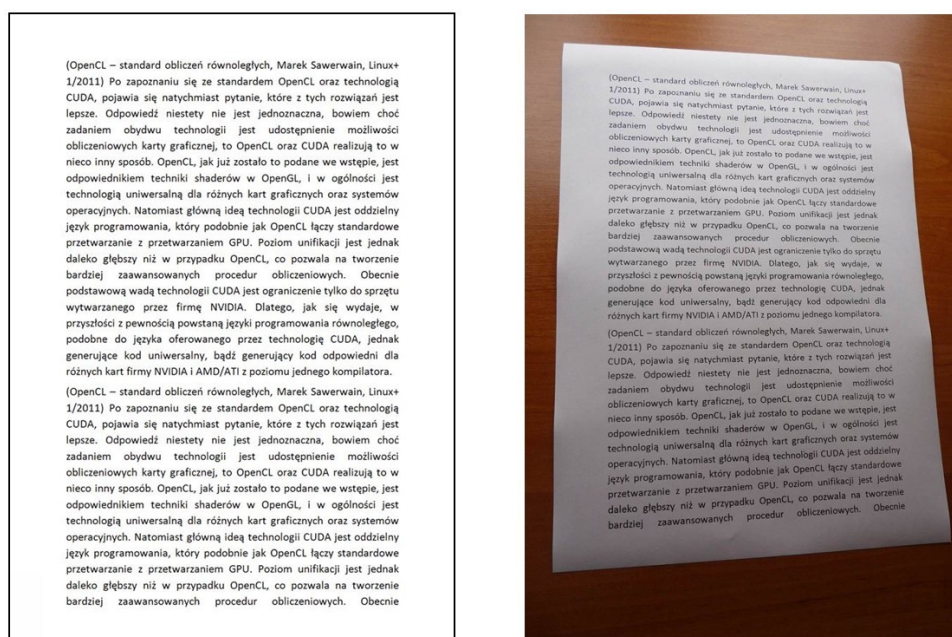
ficznego. Przy użyciu aparatu z matrycą ok. 8 megapikseli można przy kopiowanym formacie dokumentu A4 osiągnąć rozdzielczość 300 dpi, wystarczającą do dobrej jakości wydruku, przeglądania treści i po dokonaniu korekcji – automatycznego rozpoznawania tekstu. W przypadku zastosowania aparatu fotograficznego występują jednak dwa podstawowe problemy: nierównomierność oświetlenia oraz zniekształcenia geometryczne obrazu. Innymi czynnikami wpływającymi na obniżenie jakości obrazu dokumentu mogą być cienie oraz zagięcie dokumentu, np. w przypadku wykonywania zdjęć książek [1], wady układu optycznego, zbyt mała głębia ostrości.

Problem poprawiania jakości obrazów dokumentów jest rozwiązywany na wiele sposobów, zależnie od charakteru fotografowanych obiektów i spodziewanych zniekształceń. Spotykane są problemy bardziej złożone, np. korekcji dokumentów z zagnieceniami [2]. Tutaj jednak poza płaskim zdjęciem wymagane jest dodatkowe skanowanie kształtu dokumentu skanerem 3D; model trójwymiarowy pozwala na modelowanie oświetlenia i zniekształceń geometrycznych oraz ich korekcję. Podobną koncepcję opisano w [4]. Korekcji mogą też wymagać zdjęcia innych obiektów, np. terenu wykonywane z dużej wysokości. Problem opisano w [3]. Na podstawie nierównomiernie oświetlonego zdjęcia tworzona jest bitmapa – obraz referencyjny, opisujący tę nierównomierność przez zastosowanie filtracji dolnoprzepustowej (rozmycia Gaussa). Uzyskany obraz referencyjny jest odejmowany od oryginalnego obrazu, co daje efekt podobny do filtracji górnoprzepustowej, ze zmniejszeniem kontrastu jako dodatkowym skutkiem. Korekcja zniekształceń jest też konieczna w procesie rozpoznawania tekstu. Pomocne jest założenie, że korygowany obraz zawiera tekst drukowany w prostych, równoodległych wierszach. Metodę korekcji tego typu opisano w [5].

Jednym z początkowych etapów korekcji jest korekcja zniekształceń wprowadzanych przez układ optyczny aparatu fotograficznego. W przypadku cyfrowych aparatów fotograficznych korekcja tego typu może być jednak wykonywana przez oprogramowanie samego aparatu. Tak korygowane są np. zniekształcenia wprowadzane przez obiektyw, szczególnie przy krótkich ogniskowych (dystorsja beczkowa). Korekcja musi uwzględniać zależność zniekształceń od ustawionej ogniskowej. W badanych kompaktowych aparatach cyfrowych: Canon PowerShot S90 z obiektywem o ogniskowej 28 – 105 mm (w przeliczeniu na format 35 mm), Sanyo Xacti HD2000 z ob. 38 – 380 mm, Panasonic Lumix TZ5 z ob. 28 – 280 mm, korekcja zniekształceń wprowadzanych przez obiektyw była wystarczająco skuteczna, aby przyjąć, że obrazami prostych krawędzi są proste odcinki. Dla aparatu Canon PowerShot S90 porównano obraz bezpośrednio otrzymany z matrycy (RAW) z przetworzonym do wynikowego JPG. Obraz RAW wykazuje dystorsję beczkową, JPG – nie.

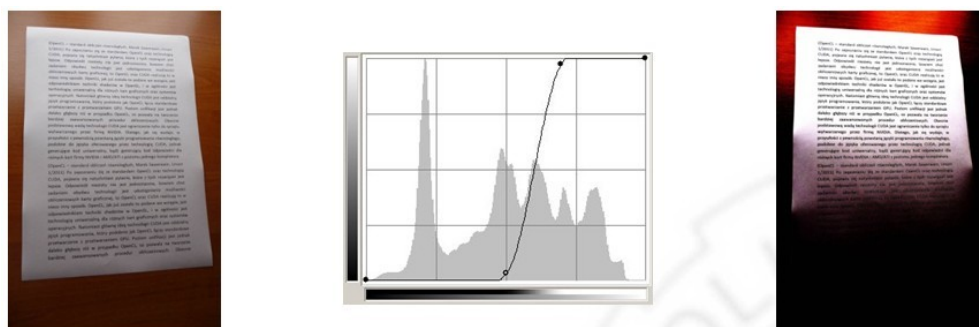
## 2. Korekcja oświetlenia

Założono, że po korekcji oświetlenia jasne (białe) obszary obrazu dokumentu powinny mieć tę samą jasność w obrębie całego dokumentu. Po operacji progowania może to być maksymalny poziom jasności. Na rys. 1 przedstawiono przykładowy dokument tekstowy oraz jego zdjęcie z nierównomiernym oświetleniem i zniekształceniem perspektywicznym. Poza nierównomiernym oświetleniem zmniejszony jest też kontrast. Uzyskany w wyniku korekcji obraz powinien być zbliżony do oryginalnego.



Rys. 1. Przykład dokumentu tekstowego i jego zdjęcia  
Fig. 1. An example of a text document and its image

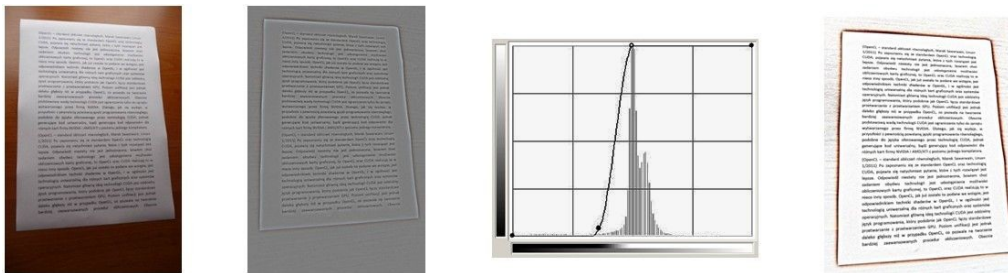
Z uwagi na nierównomierność oświetlenia proste operacje na jasności typu progowanie, zmiana jasności, kontrastu nie dają oczekiwanego efektu. Niezależnie od wyboru progu rozdzielającego poziom jasności obrazu białych i czarnych obszarów dokumentu, niektóre obszary są błędnie klasyfikowane, jak pokazano na rys. 2. Na rysunku przedstawiono oryginalny obraz, jego histogram i krzywą odwzorowania oraz wynik przekształcenia.



Rys. 2. Prosta korekcja oświetlenia  
Fig. 2. A simple lighting correction

W przypadku idealnego obrazu histogram powinien wykazywać dwa maksima w okolicach wartości minimalnej (czerni) i maksymalnej (biel); tutaj wartości są rozproszone i niezależnie od wyboru progu uzyskujemy błędny wynik podobny do przedstawionego w prawej części rys. 2. Konieczne jest więc użycie innych metod.

Pierwszą metodą korekcji jasności jest dokonanie filtracji górnoprzepustowej, a następnie korekcji kontrastu i opcjonalnie progowania. Wynik takiego procesu przedstawia rys.3. Pokazano kolejno oryginalne zdjęcie, wynik filtracji górnoprzepustowej, histogram obrazu po filtracji i krzywą odwzorowania oraz wynik.



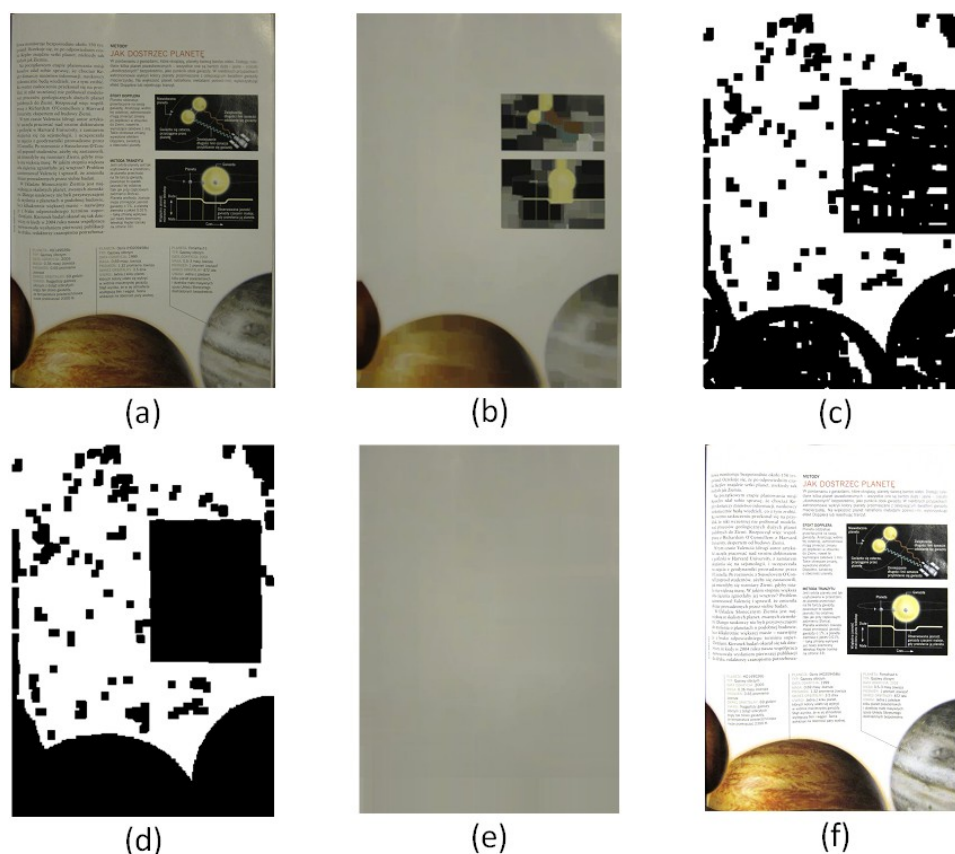
Rys. 3. Korekcja oświetlenia przez filtrację górnoprzepustową i progowanie  
Fig. 3. Correction of lighting by high-pass filtering and thresholding

Filtracja górnoprzepustowa jest dokonywana przez zastosowanie filtru rozmywającego Gaussa, a następnie odjęcie wyniku od oryginalnego obrazu i uwzględnienie składowej stałej. Promień rozmycia dla badanych zdjęć wynosił ok. 50 pikseli. Tak duży promień pozwala na uśrednienie jasności obszaru o wielkości porównywalnej lub większej od rozmiaru liter tekstu.

Operacja progowania może być opcjonalnie wykonana po korekcji kontrastu. W wyniku progowania uzyskujemy czarno-biały obraz bez odcieni szarości. Jest to pożądane dla obrazów ze względnie dużą czcionką tekstu. Dla mniejszych rozmiarów progowanie może negatywnie wpłynąć na jakość zobrazowania tekstu na skutek usunięcia szarych pikseli wygładzających krawędzie (antialiasing).

Dla dokumentów zawierających zdjęcia i grafikę powyższa metoda może doprowadzić do utraty treści tego rodzaju obiektów. Jest to skutkiem działania filtra górnoprzepustowego, który usuwa składową stałą i składowe wolnozmiennie z obrazu, podczas gdy zdjęcia i rysunki mogą takie składowe zawierać. Efektem jest pozostawienie tylko konturów kształtów przedstawionych na zdjęciu. Dlatego zastosowano w tym przypadku inną metodę. Kolejne jej etapy przedstawiono na rys. 4.

Założono, że dokument ma postać tekstu z niewielkimi wstawkami graficznymi, a tło jest białe jak na rys. 4a. Dla innych postaci dokumentu wynik korekcji oświetlenia może być niezgodny z oczekiwaniami. Celem pierwszego etapu jest ustalenie zmienności natężenia oświetlenia obszaru kartki i wygenerowanie hipotetycznego obrazu białej kartki, oświetlonej tak samo jak analizowany dokument (rys. 4e). Obraz ten jest nazywany referencyjnym.



Rys. 4. Korekcja oświetlenia dokumentów zawierających tekst i grafikę  
Fig. 4. Lighting correction of documents containing text and graphics

Obraz referencyjny jest tworzony w następujący sposób. Obraz wejściowy jest pomniejszany do rozmiaru ok. 0,1 megapiksela. Następnie są usuwane z obrazu wszystkie elementy, tak aby pozostało tylko tło – obraz referencyjny. Dla elementów o niewielkich rozmiarach (tekstu, cienkich liniach) skutecznym sposobem ich usunięcia jest zastosowanie operacji erozji, zmniejszającej powierzchnię czarnych obszarów przez usuwanie brzegowych pikseli. Zastosowano element strukturalny o rozmiarze 3x3. Erozja może jednak spowodować rozjaśnienie tła obrazu w obszarach nierównomiernie oświetlonych. W celu kompensacji rozjaśnienia po erozji jest wykonywana przeciwnie działająca operacja dylatacji. Operacje są powtarzane kilkakrotnie. Wynik tego etapu przedstawia rys. 4b. Na obrazie pozostały jedynie duże obiekty.

Obiekty te (np. zdjęcia) nie zostały usunięte ze względu na rozmiar znacznie większy od rozmiaru elementu strukturalnego. Usuwanie takich elementów jest więc realizowane inną metodą: przez wykrycie takich obszarów i wypełnienie ich kolorem tła z obrzeża. Przy zastosowaniu filtra wykrywającego krawędzie i progowania z niską wartością progu obraz jest dzielony na obszary o podobnej jasności. Na wynik nie mają wpływu łagodne zmiany jasności. Rysunek 4c przedstawia wynik takiego podziału. Biały kolor oznacza małą lokalną zmianę jasności, czarny – dużą. W niektórych miejscach duża lokalna zmiana jasności tła spowodowała powstanie bloków uznanych błędnie za nienależące do tła. Również niektóre

miejsca zdjęć zostały uznane za tło (rys. 4c). Maski dzieli kartkę na wiele białych obszarów. Największy z nich zostaje uznany za tło; pozwala to na wypełnienie brakujących fragmentów maski (rys. 4d). Obrazy z rys. 4b i rys. 4d są podstawą wyznaczenia obrazu referencyjnego z rys. 4e. Czarne obszary na 4d oznaczają, że jasność odpowiednich pikseli obrazu referencyjnego musi zostać wyznaczona przez interpolację jasności na ich brzegach. Obraz referencyjny jest skalowany do rozmiaru obrazu wejściowego.

Tak uzyskany obraz referencyjny jest podstawą dokonania ostatecznej korekcji oświetlenia. Składowe R, G, B każdego piksela obrazu wejściowego są mnożone przez  $255/\text{ref}(x,y)$ , gdzie  $\text{ref}(x,y)$  jest jasnością piksela  $(x,y)$  obrazu referencyjnego.

### 3. Korekcja geometrii

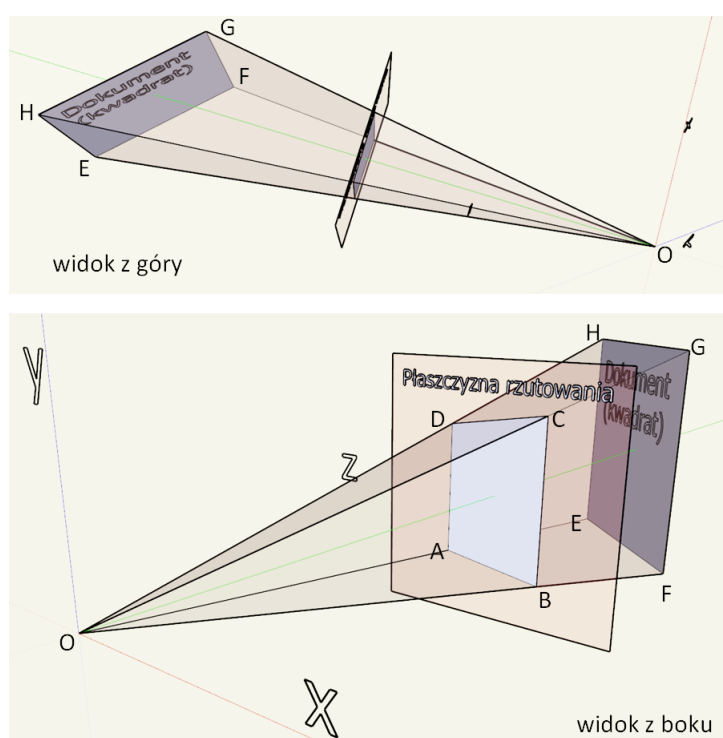
Kolejny etap przetwarzania dokonuje korekcji zniekształceń geometrycznych powodowanych przez perspektywę. Podobnie jak poprzedni powinien działać w sposób automatyczny, bez ingerencji użytkownika. W ogólnym przypadku, kiedy nieznan jest rzeczywisty kształt dokumentu, taka operacja nie jest możliwa – wiele różnych konfiguracji kształtu dokumentu i orientacji aparatu da taki sam obraz wynikowy. Dlatego założono, że dokument ma kształt prostokąta, w całości widocznego w kadrze. Krawędziami obrazu również są odcinki prostych. Korekcji podlega zniekształcenie wynikające z perspektywy, przedstawione na rys. 5, widoczne również na rys. 2 i 3. Przykładowy dokument z rys. 4 ma kształt kwadratu EFGH, natomiast jego obraz jest czworokątem ABCD o nierównych i nierównoległych bokach.

Korekcja przeprowadzana jest w następujący sposób:

- zmniejszenie obrazu,
- filtrowanie obrazu uwydatniające krawędzie,
- binaryzacja wyniku – uzyskanie obrazu krawędzi,
- transformata Hougha – odszukanie prostych o różnych nachyleniach i położeniach,
- wybór odcinków – potencjalnych brzegów obrazu dokumentu,
- wybór konturu dokumentu i wyznaczenie współrzędnych wierzchołków,
- przekształcenie odwrotne do perspektywicznego.

Przy założeniu prostokątnego kształtu dokumentu wystarczy odnaleźć na obrazie kształt czworokąta wypukłego, który jest krawędzią dokumentu, wyznaczyć hipotetyczne współrzędne wierzchołków prostokąta w przestrzeni 3D, dokonać mapowania punktów zdjęcia na ten prostokąt i ponownie rzutować je na inną płaszczyznę, równoległą do tego prostokąta, przez rzut prostokątny. Najtrudniejszym elementem tego procesu jest odnalezienie obrazów wierzchołków prostokąta zawierającego dokument.





Rys. 5. Podlegające korekcji zniekształcenie obrazu wynikające z perspektywy  
 Fig. 5. Perspective distortion of an image to be corrected

Początkowe etapy procesu nie wymagają komentarza. Zastosowana transformata Hougha działa na obrazie binarnym, zawierającym uwidocznione krawędzie (umownie: czarne piksele oznaczają krawędzie). Transformata Hougha pozwala wykryć na obrazie zbiory pikseli leżących na jednej prostej. Jest odporna na niewielkie błędy pozycji pikseli, dzięki czemu wykryte zostaną również proste nieco zniekształcone. Dla wszystkich możliwych orientacji i położenia prostych jest określana liczba czarnych pikseli obrazu leżących na tej prostej. Wynikiem jest dwuwymiarowa tablica tzw. akumulatorów, zawierająca liczby czarnych pikseli wzdłuż odpowiednich prostych. Dwa wymiary tablicy odpowiadają liczbie parametrów opisujących dowolną prostą. Duże wartości w tablicy (maksima lokalne) odpowiadają wykrytym odcinkom. Oczekujemy dokładnie czterech prostych zawierających krawędzie, ale transformata może zwrócić ich więcej. W takim przypadku jest konieczny wybór czterech połączonych odcinków. Wybierane są one na podstawie oceny kątów (tworzony czworokąt musi być wypukły) i obejmowanego obszaru (jak największy).

Po ustaleniu współrzędnych czterech wierzchołków obrazu kartki są wyznaczane hipotetyczne współrzędne wierzchołków kartki w trójwymiarowej przestrzeni. Oznaczmy jak na rys. 5 położenie obserwatora przez  $O$  (jest to też początek układu współrzędnych), wierzchołki prostokątnego dokumentu przez  $E, F, G, H$ , wierzchołki obrazu dokumentu na płaszczyźnie prostopadłej do osi  $OZ$  przez  $A, B, C, D$ . Płaszczyzna rzutowania znajduje się w odległości  $d$  od początku układu współrzędnych. Oznaczmy też wektory:  $\vec{v}_1 = \vec{OA}$ ,  $\vec{v}_2 = \vec{OB}$ ,  $\vec{v}_3 = \vec{OC}$ ,  $\vec{v}_4 = \vec{OD}$ ,  $\vec{u}_1 = \vec{OE}$ ,  $\vec{u}_2 = \vec{OF}$ ,  $\vec{u}_3 = \vec{OG}$ ,  $\vec{u}_4 = \vec{OH}$ . Wektory  $\vec{v}_i$  mają ten sam zwrot

i kierunek jak  $\vec{u}_i$ , różnią się długością. Przyjmijmy  $\vec{u}_i = t_i \cdot \vec{v}_i$  dla  $i = 1, \dots, 4$ . Znamy tylko położenia punktów O, A, B, C, D; szukamy współrzędnych E, F, G, H. Zakładamy, że punkty E, F, G, H leżą w jednej płaszczyźnie i są wierzchołkami prostokąta. Spełnione są więc zależności (1) i (2):

$$\vec{FE} = \vec{GH} \equiv (\vec{u}_1 - \vec{u}_2) = (\vec{u}_4 - \vec{u}_3), \quad (1)$$

$$\vec{EF} \cdot \vec{FG} = (\vec{u}_2 - \vec{u}_1) \cdot (\vec{u}_3 - \vec{u}_2) = 0. \quad (2)$$

Warunek (1) można zapisać dla poszczególnych współrzędnych jako układ równań (3):

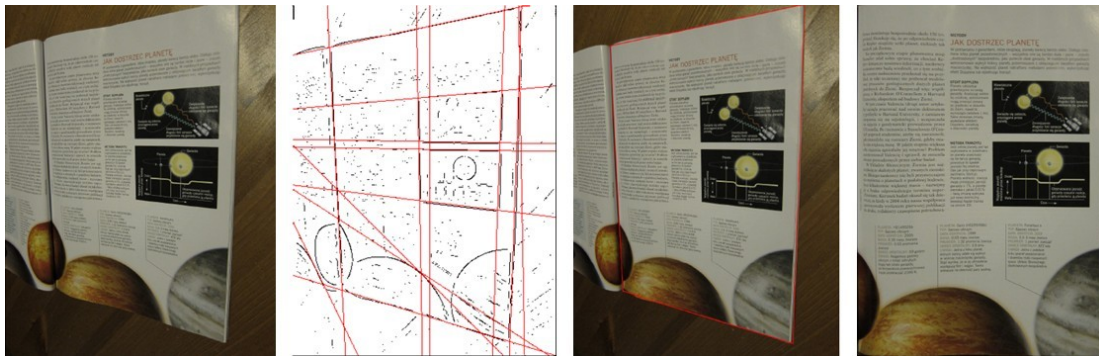
$$\begin{cases} x_E - x_F = x_H - x_G \\ y_E - y_F = y_H - y_G \\ z_E - z_F = z_H - z_G \end{cases} \quad (3)$$

Ponieważ  $\vec{u}_i = t_i \cdot \vec{v}_i$ , można zapisać te równania jako (4):

$$\begin{cases} t_1 x_A - t_2 x_B = t_4 x_D - t_3 x_C \\ t_1 y_A - t_2 y_B = t_4 y_D - t_3 y_C \\ t_1 z_A - t_2 z_B = t_4 z_D - t_3 z_C \end{cases} \quad (4)$$

W ostatnim równaniu wszystkie współrzędne z punktów A, B, C, D są równe  $d$ ; dodatkowo, możemy jedną z wartości  $t_i$  przyjąć jako dowolną, np.  $t_1 = 1$  (A=E), ponieważ nie da się jednoznacznie określić odległości dokumentu od płaszczyzny rzutowania. Układ równań przyjmie postać układu równań liniowych z trzema niewiadomymi  $t_2, t_3, t_4$  (5):

$$\begin{cases} x_A - t_2 x_B + t_3 x_C - t_4 x_D = 0 \\ y_A - t_2 y_B + t_3 y_C - t_4 y_D = 0 \\ 1 - t_2 + t_3 - t_4 = 0 \end{cases} \quad (5)$$



Rys. 6. Korekcja zniekształceń geometrycznych wynikających z perspektywy

Fig. 6. Correction of geometric distortions resulting from the perspective

Warunek (2) można zapisać w postaci równania (6):

$$\begin{aligned} & (t_2 x_B - x_A)(t_3 x_C - t_2 x_B) + \\ & (t_2 y_B - y_A)(t_3 y_C - t_2 y_B) + \\ & d^2(t_2 - 1)(t_3 - t_2) = 0 \end{aligned} \quad (6)$$

Po wyznaczeniu  $t_2, t_3, t_4, d$  z (5) i (6) można wyznaczyć współrzędne wierzchołków E, F, G, H w przestrzeni trójwymiarowej i rzut prostokątny dokumentu na płaszczyznę do niego równoległą, co kończy etap korekcji perspektywy. Na rys. 6 przedstawiono kolejne etapy korekcji geometrii przykładowego obrazu.



## 4. Implementacja i wnioski

Opisane metody zostały zaimplementowane w języku C jako wtyczka do programu GIMP. To rozwiązanie pozwala na skorzystanie z rozwiniętego oprogramowania przetwarzającego obrazy, wyposażonego zarówno w podstawowe funkcje (np. import obrazów w różnych formatach, podgląd, edycja), jak i zaawansowane (obsługa wielu warstw, filtry itp.). Dołączenie opisanej wtyczki rozszerza możliwości programu. Integracja wtyczki z GIMP-em odbywa się dzięki bibliotece libgimp zawierającej funkcje i makra pozwalające zarejestrować główną funkcję wtyczki w bazie procedur GIMP-a. Wtyczka zawiera interfejs użytkownika pozwalający na konfigurację podstawowych parametrów zastosowanych algorytmów.

Przeprowadzone eksperymenty wykazały skuteczne działanie opisanych metod przy przetwarzaniu fotografowanych dokumentów oraz innych obiektów o założonym kształcie. Uzyskanie dobrych efektów jest uwarunkowane spełnieniem założeń dotyczących kształtu dokumentu oraz odpowiednimi warunkami wykonania zdjęcia, pozwalającymi na automatyczne wykrycie krawędzi dokumentu i częściową analizę jego treści. W przypadku wystąpienia zniekształceń o innym charakterze nie zostaną one w pełni skorygowane.

## BIBLIOGRAFIA

1. Zhang Z., Lim Tan C.: Restoration of images scanned from thick bound documents. International Conference on Image Processing, 2001, Vol. 1, s. 1074÷1077.
2. Sun M., Yang R., Yun L., Landon G., Seales B., Brown, M.S.: Geometric and photometric restoration of distorted documents. International Conference on Computer Vision, 2005, Vol. 2, s. 1117÷1123.
3. Wang M., Pan J., Chen S., Li H.: A method of removing the uneven illumination phenomenon for optical remote sensing image. Geoscience and Remote Sensing Symposium, 2005, Vol. 5, s. 3243÷3246.
4. Landon G.V., Yun L., Seales W.B.: Towards Automatic Photometric Correction of Casually Illuminated Documents. Computer Vision and Pattern Recognition, 2007.
5. Junjuan L., Tan Guoxin T.: An Efficient Algorithm for Skew-Correction of Document Image Based on Cyclostyle Matching. Computer Science and Software Engineering, 2008, s. 1267÷1270.

## **Abstract**

Creating a digital archive of documents available in a paper form has many reasons, among them more convenient access, lower volume, easy indexing and searching after performing OCR. The standard method for the conversion of printed documents into an electronic form is scanning. With proper selection of equipment and scanning parameters the quality of the copy is comparable with the original. However, obtaining such a good quality is time consuming and requires special preparation of source materials. A much faster method is to use a digital camera. Pictures can be taken up every two seconds with 300 dpi resolution for A4-formatted documents – sufficient for a good quality print. Unfortunately, there are problems with the quality of such copies. These include uneven illumination, geometric distortion of the image, shadows etc.

At least two of these elements can be corrected automatically. The developed methods assume, that the document is a flat, rectangular sheet of paper. It should also be fully visible, as one of the steps of the correction algorithm is to detect edges of a rectangular sheet. Unevenness of illumination can be corrected by performing a high-pass image filtering. For documents containing text and images (photos, graphics), this method can cause abnormal lightening of internal fragments of illustration. Therefore, a method which excludes these areas from filtering is presented. The second element is geometry correction, for images of documents meeting the conditions specified earlier. The described method corrects distortions resulting from perspective. In the first step the location of the document edges is determined using the Hough transform. Then, hypothetical coordinates of the vertices of the document in the 3D space are determined and a reverse transformation is performed. These algorithms are implemented as a plug-in for the GIMP graphics program.

## **Adresy**

Piotr FABIAN: Politechnika Śląska, Instytut Informatyki, ul. Akademicka 16,  
44-100 Gliwice, Polska, Piotr.Fabian@polsl.pl.

Jacek OLEARCZYK: Jacek.Olearczyk@gmail.com.