

*MAREK KANNCHEN*

Państwowa Wyższa Szkoła Zawodowa w Gorzowie Wielkopolskim

**ZASTOSOWANIE METOD ANALIZY OBRAZU DO  
WSPOMAGANIA OSÓB NIEWIDOMYCH NA UCZELNI  
WYŻSZEJ**

**Streszczenie**

W artykule zostanie przedstawiony przegląd metod analizy obrazów pod kątem możliwości wykorzystania w systemie wspomagającym osoby niewidome w budynkach uczelni. System taki może przekazywać niewidomemu użyteczne dla niego informacje. Pozyskanie tego typu informacji wymaga m.in. analizy obrazów z kamer. W tego typu analizie kluczową rolę odgrywać będą metody dobrane w taki sposób, aby zapewnić jak najmniejsze zużycie energii przez część systemu, w który niewidomy może zostać wyposażony. W celu uproszczenia metod można wykorzystać czujniki dostarczające informacje, które w innym przypadku musiałyby zostać pozyskane poprzez dodatkową analizę obrazu. Wynikiem przeglądu będzie wskazanie metod najwłaściwszych do wykorzystania w systemie.

**Wprowadzenie**

Miniaturyzacja urządzeń elektronicznych daje nowe możliwości w ich zastosowaniu. Urządzenia te są coraz mniejsze, potrzebują coraz mniej energii,

więc łatwiejsze jest ich zastosowanie w różnych dziedzinach życia. W artykule dokonano przeglądu metod analizy obrazów pod kątem wykorzystania w systemie wspomagającym osoby niewidome w budynkach uczelni. Właśnie dzięki miniaturyzacji urządzeń elektronicznych system taki dałby niewidomym możliwość pozyskiwania i przetwarzania obrazów na ich potrzeby, a przy tym jego wymiary nie byłyby uciążliwe w użytkowaniu dla osoby nim się posługującej.

W skład systemu, który wspomagałby osoby niewidome na uczelni, wchodzić powinny następujące elementy:

1. Mikrokamera. Pozwoli ona na pozyskanie materiału (obrazów) do późniejszej analizy.
2. Nośnik do przechowywania obrazów. Najodpowiedniejszym nośnikiem byłaby karta pamięci – ze względu na jej małe rozmiary.
3. Mikrokontroler służący do wykonywania operacji na pozyskanych obrazach.
4. Syntezator mowy. Umożliwi zamianę informacji pozyskanych w postaci obrazów na komunikaty słowne.

Dla pozyskania lepszej jakości materiału do analizy można by zastosować w systemie również sonar (czujnik mierzący odległość), akcelerometr, kompas – magnetometr (określenie położenia kamery względem pionu). Pozwoliłoby to na wstępną korekcję poszczególnych klatek obrazu pozyskanych z kamery.

Celem niniejszego artykułu jest przedstawienie tych metod analizy obrazu, które mogą być użyte w takim systemie w sposób optymalny, wykorzystując jego budowę, a zarazem nie obciążając zasobów, by system mógł dokonywać analizy i podawać odpowiednie komunikaty osobie korzystającej z systemu w czasie zbliżonym do rzeczywistego.

W artykule tym przedstawiono dwa przypadki wykorzystania metod analizy obrazów pod kątem wspomagania osób niewidomych w budynkach uczelni wyższej. W rozdziale pierwszym zajęto się przypadkiem, gdy osoba niewidoma potrzebuje odczytać informacje przekazane w postaci tekstu na tablicy informacyjnej. W rozdziale drugim natomiast opisano sytuację, gdy osoba posługująca się systemem znajdzie się w pomieszczeniu, w którym znajdują się przedmioty niebędące stałym elementem wyposażenia budynku, to znaczy mogące zmieniać swoje położenie, a tym samym stwarzać zagrożenie dla osoby niewidomej. System w takim przypadku wykorzystany zostałby do ostrzeżenia przed niebezpieczeństwem grożącym osobom nim wspomaganym ze strony obiektów „ruchomych”.

W procesie analizy obrazów istotny jest dobór rozdzielczości obrazu. Rozdzielczość jest miarą zdolności rozpoznawania szczegółów obrazu. Wybór właściwej rozdzielczości jest kompromisem pomiędzy szczegółowością obrazu a czasem przetwarzania i objętością obrazu. Obrazy o dużej rozdzielczości zawierają więcej szczegółów, można z nich odczytać więcej informacji, natomiast większa rozdzielczość obrazu powoduje wydłużenie czasu jego analizy. Zatem najczęściej pozyskuje się obraz o dużej rozdzielczości, by w trakcie początkowych etapów jego obróbki dojść do rozdzielczości praktycznej dla procesów analizy. Najczęstsze rozdzielczości wykorzystywane podczas analizy obrazów są całkowitymi potęgami dwójki od  $256 \times 256$  do  $4096 \times 4096$ <sup>1</sup>.

W przypadku gdy materiał (obraz) ma być dostarczony do analizy przez mikrokamerę, w którą wyposażony będzie system, należy również uwzględnić rozdzielczość tej mikrokamery. Zatem w artykule analizie poddano obrazy o rozdzielczości  $512 \times 512$  pikseli. Rozdzielczość ta jest podobna do tej, którą można pozyskać w mikrokamerach, których zakup nie będzie wiązał się z wysokimi kosztami.

## 1. Czytanie tekstu z tablicy

Czytanie tekstu jest jednym z problemów, z jakimi spotyka się osoba niewidoma studiująca na uczelni wyższej. Z informacjami podanymi za pomocą tekstu człowiek spotyka się na co dzień. Również w czasie studiów studentom podawane są różne wiadomości tekstowe na tablicach informacyjnych – począwszy od planu zajęć, komunikatów władz uczelni, na wynikach z zaliczeń i kolokwium skończywszy. Student niewidomy, wyposażony w system rozpoznawania i przetwarzania tekstów na komunikaty głosowe, na pewno czułby się bardziej komfortowo i niezależnie, nie musząc prosić swoich kolegów o pomoc w tego typu sytuacjach.

Proces przekształcania tekstu otrzymanego z mikrokamery w postaci obrazu na komunikat tekstowy jest procesem wieloetapowym.

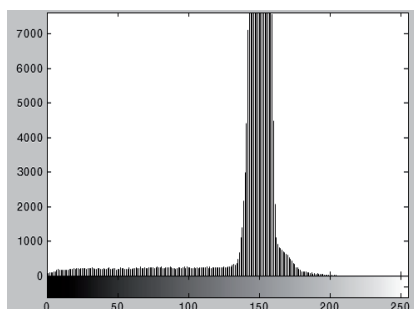
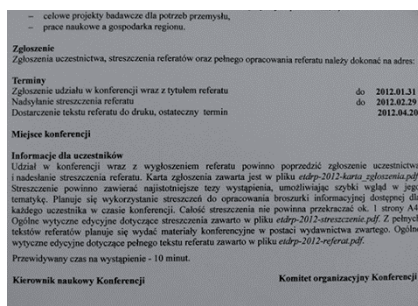
Pierwszym etapem do przetworzenia tekstu z obrazu na komunikat głosowy jest wykrycie tekstu w analizowanym obrazie. Możemy użyć do tego

---

<sup>1</sup> R. Tadeusiewicz, P. Korohoda, *Komputerowa analiza i przetwarzanie obrazów*, Wydawnictwo Fundacji Postępu Telekomunikacji, Kraków 1997.

celu histogramu obrazu poddawanego analizie. Przedstawia on, w formie graficznej lub tablicy, rozkład liczbowy występowania w obrazie cyfrowym poszczególnych poziomów jasności<sup>2</sup>.

Gdy analizowany obraz zawiera bloki tekstu jak na rysunku 1a, jego histogram znacznie różni się od histogramu obrazu przedstawiającego jakieś obiekty.

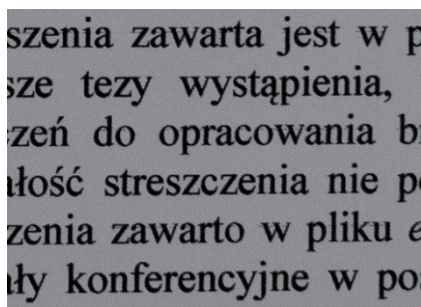


Rys. 1a. Obraz zawierający bloki tekstu Rys. 1b. Obraz zawierający histogram

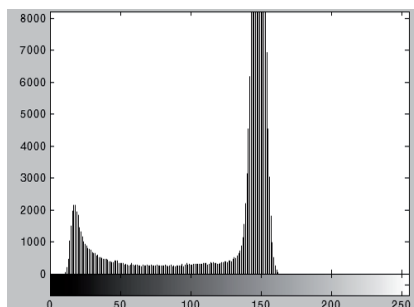
Z histogramu na rys. 1b odczytać możemy, że najliczniej w obrazie występują poziomy szarości z przedziału 140–160. Wynika to ze słabego doświetlenia analizowanego obrazu. W prawidłowo doświetlonym obrazie najliczniej występowałyby wartości z przedziału około 250, odpowiada jasnym pikselom. Natomiast dla pikseli ciemnych zbliżonych do czerni rozkład jest w miarę równomierny – ma tę samą wielkość dla wartości zaczynających się około 10 i kończących się w okolicach wartości 140.

Analizy obrazu możemy dokonać w dwóch etapach. Najpierw badać możemy cały obraz i wtedy histogram mógłby wyglądem przypominać histogram z rysunku 1b. Aby upewnić się, że obraz na pewno zawiera tekst, możemy w drugim etapie poddać analizie fragment obrazu po przybliżeniu. Wtedy jego histogram może przypominać wyglądem histogram z rysunku 2b. Charakterystyczny dla takiego przybliżenia jest wzrost wartości dla pikseli zbliżonych do czerni dla wartości od 10 do 20.

<sup>2</sup> W. Malina, M. Smiatacz, *Cyfrowe przetwarzanie obrazów*, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2008; A. Stateczny, *Metody nawigacji porównawczej*, Gdańskie Towarzystwo Naukowe, Gdynia 2004.



Rys. 2a. Zbliżenie obrazu zawierającego bloki tekstu



Rys. 2b Histogram zblizenia obrazu zawierajacego bloki tekstu

Po wykryciu tekstu powinniśmy dokonać zamiany obrazu kolorowego na dwupoziomowy. Konwersja taka przebiega w kilku krokach. Pierwszym z nich jest konwersja do obrazu czarno-białego.

Następnie należy dokonać konwersji obrazu otrzymanego w poprzednim kroku do obrazu dwupoziomowego. Zmiana ta odbywa się poprzez metody progowania.

Progowanie polega na porównywaniu wartości każdego punktu obrazu z zadaną wartością progową. Jeśli odpowiednio dobierzemy wartość progową, możemy w obrazie wyodrębnić obszary danego typu. W najprostszej metodzie stosuje się jedną wartość progę i jest ona realizowana według zależności<sup>3</sup>:

$$J_w(x, y) = \begin{cases} 1, & J(x, y) > t \\ 0, & J(x, y) \leq t \end{cases} \quad (1.1)$$

gdzie  $t$  – próg binaryzacji.

Powyższa operacja oznacza zamianę obrazu źródłowego w odcieniach szarości na obraz binarny. Punkty, dla których  $J_w(x, y) = 1$ , są to w nowym obrazie zwykle punkty obiektu (przypisuje się im kolor czarny). Pozostałe punkty obrazu tworzą tło (przypisuje się im kolor biały). Jeśli zamienimy przyporządkowywane pikselom wartości, w wyniku progowania otrzymamy negację obrazu binarnego.

<sup>3</sup> W. Malina, M. Smiatacz, dz. cyt.

Binaryzacja obrazu dokumentu tekstowego jest ważnym przykładem powyższej operacji. Na obrazie takim występują najczęściej dwie klasy obiektów – znaki (litery, cyfry) i tło. Znaki to obiekty, które chcemy wydzielić z tła, by następnie poddać operacji rozpoznawania. Pozostałe elementy obrazu tworzące tło – usuwamy. Jeżeli tło ma dużo zakłóceń, operacja ta jest bardziej skomplikowana.

Obraz  $I$ , który chcemy poddać progowaniu, można wstępnie wygładzić. Nie jest to operacja niezbędna, ale w praktyce często stosowana, ponieważ wygładzony obraz łatwiej jest poddać progowaniu. Następnie wyznacza się histogram, konstruuje funkcję kryterialną  $\eta(t)$  i wyznacza pierwszy próg  $t_1$ . Próg ten dzieli obraz na dwie klasy  $c_0$  (tło) i  $c_1$  (obiekt). W kolejnych krokach rekursji  $n = 2, 3, \dots, p$  wyznacza się histogramy poziomów szarości i progi dla obrazu ograniczonego. Rekursję powtarza się  $p$  razy, aż do spełnienia kryterium końca. Wtedy obraz źródłowy  $I$  możemy przedstawić jako sumę obrazów posegmentowanych<sup>4</sup>:

$$I = c_0 \vee c_1 \vee \dots \vee I_p \quad (1.2)$$

gdzie  $I_p$  – obiekt pozostały po wielokrotnym progowaniu.

Kolejnym krokiem jest konturyzacja pozwalająca wykryć krawędzie, narożniki obiektów w analizowanym obrazie. Stosuje się w tym celu filtry. Jest to tak zwana filtracja konwolucyjna.

Jednym z takich filtrów jest filtr górnoprzepustowy wykrywający krawędzie, zwany laplasjanem. Laplasjan, dając dobre efekty w prostych zadaniach, pozwala na wykrywanie i podkreślanie na obrazie wszelkich krawędzi i konturów obiektów bez względu na to, pod jakim kątem one przebiegają. W przetwarzaniu obrazów natomiast laplasjanem nazywa się wynik konwolucji obrazu maską  $3 \times 3^5$  lub maską nieco bardziej rozbudowaną<sup>6</sup>.

<sup>4</sup> Tamże.

<sup>5</sup> R. Tadeusiewicz, P. Korohoda, *Komputerowa analiza i przetwarzanie obrazów*, Wydawnictwo Fundacji Postępu Telekomunikacji, Kraków 1997; G. Ritter, J. Wilson, *Handbook of Computer Vision Algorithms in Image Algebra*, CRC Press, Boca Raton 2001; B. Jähne, *Practical Handbook on Image Processing for Scientific and Technical Applications*, CRC Press, 2004.

<sup>6</sup> R. Tadeusiewicz, P. Korohoda, dz. cyt.; C. Russ, *The image processing handbook*, CRC Press, Boca Raton 2007.

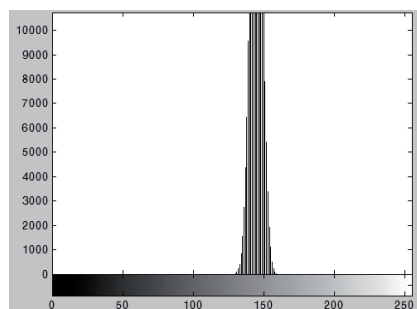
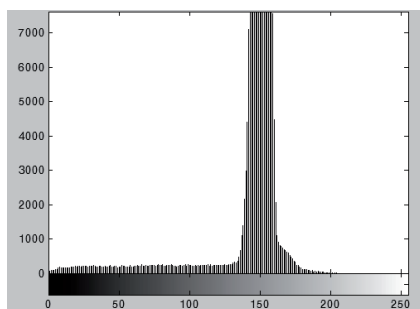
Możemy też wyznaczyć laplasjan przy wykorzystaniu macierzy konwolucji większej niż  $3 \times 3$ , na przykład maską  $5 \times 5$ <sup>7</sup>.

Tak przygotowany obraz możemy poddać operacji ścieniania. Ścienianie to wspólna nazwa dla podzbioru przekształceń morfologicznych. Jednym z przekształceń, w których stosuje się ścienianie, jest szkieletyzacja. Jest to operacja pozwalająca wyodrębnić osiowe punkty (szkielety) figur w analizowanym obrazie.

Szkielet figury jest zbiorem wszystkich punktów, które są równoległe od co najmniej dwóch punktów należących do brzegu. Jest znacznie mniejszy od niej, a w pełni odzwierciedla jej podstawowe topologiczne własności.

Szkieletyzację możemy realizować jako ścienianie z elementem strukturalnym w postaci macierzy  $3 \times 3$ <sup>8</sup>.

Ostatnie etapy to podział tekstu na linie, słowa i litery. Tutaj również można dokonać analizy na podstawie histogramu, wykorzystując fakt, że histogram tekstu różni się od histogramu pustej linii, co widać na rysunku 3a i 3b.



Rys. 3a. Histogram dla obrazu z tekstem

Rys. 3b. Histogram dla pustej linii

Wykorzystując różnice między histogramem dla obrazu z tekstem a histogramem obrazu pustej linii można dokonać podziału na linie zawierające tekst. Następnie należy wyodrębnione linie, w których występuje tekst, podzielić na poszczególne słowa. Słowa te później dzielimy na pojedyncze znaki.

<sup>7</sup> R. Tadeusiewicz, P. Korohoda, dz. cyt.

<sup>8</sup> Tamże; I. Young., J. Gerbrands, L. Vliet, *Fundamentals of Image Processing*, Delft University of Technology, Delft 1998.

Tak przygotowany tekst możemy za pomocą jednej z metod rozpoznawania tekstu odczytać i zamienić na dźwięki podane poprzez system do ucha użytkownika systemu.

## 2. Wykrywanie ruchu z filmu

By umożliwić osobie niewidomej komfortowe korzystanie z budynków uczelni, można zastosować różne rozwiązania. Na przykład udźwiękowanie otoczenia, wprowadzenie do systemu planu budynku wraz z jego wyposażeniem, by osoba niewidoma miała na bieżąco informacje o otaczających ją przedmiotach w pomieszczeniu, w którym się w danej chwili znajduje. Drugie rozwiązanie wymaga jednak rozwiązania problemu ostrzegania o przedmiotach, które nie są stałym elementem wyposażenia budynku i mogą zmieniać swoje położenie, np. krzesła, ławki, kosze na śmieci itd. W tym przypadku można wykorzystać wykrywanie ruchu z filmu zarejestrowanego przez mikrokamerę znajdującą się w wyposażeniu systemu.

Pierwszym etapem analizy zarejestrowanego przez mikrokamerę filmu, jaki powinniśmy przeprowadzić na otrzymanym do analizy materiale, jest usuwanie szumu przetwornika. Złe oświetlenie otoczenia, w którym nagrywany jest obraz poddawany przez system analizie, może wprowadzać zaszumienie poszczególnych klatek obrazu i utrudniać jego późniejszą obróbkę. Dlatego przed poddaniem materiału otrzymanego z kamery analizie wykrywania ruchu należałoby obraz odszumić. Możemy tę czynność wykonać za pomocą filtracji dolnoprzepustowej albo filtracji uśredniającej.

Po wstępnej obróbce polegającej na usunięciu zakłóceń w postaci szumu należy usunąć przesunięcie klatek filmu, tak by analizowane klatki pokrywały się z sobą, to znaczy by stałe elementy wyposażenia pokrywały się na poszczególnych klatkach filmu. Do porównywania kolejno uzyskiwanych obrazów możemy użyć korelacji. Wymaga ona ciągu sekwencji minimum dwóch obrazów. Pierwszy jest traktowany jako obraz wzorca, kolejne obrazy są porównywane.

Dwuwymiarową funkcję korelacji wzajemnej można zdefiniować w następujący sposób<sup>9</sup>:

---

<sup>9</sup> R.C. Gonzalez, R.E. Woods, *Digital Image Processing Second Edition*, Prentice-Hall, Inc., New Jersey 2002.



$$f(x, y) * h(x, y) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n)h(x - m, y - n) \quad (1.3)$$

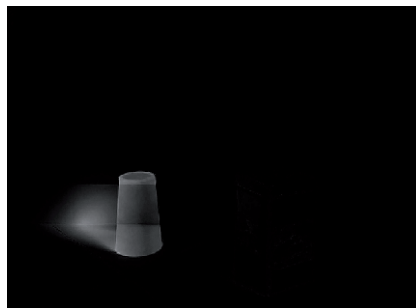
Kolejnym etapem w wykrywaniu ruchu z filmu jest sprawdzenie, czy jakiś obiekt nie przemieścił się pomiędzy kilkoma jego klatkami. Możemy w tym celu wykorzystać lokalne odchylenie standardowe.

Matematycznie lokalne odchylenie standardowe możemy przedstawić za pomocą wzoru<sup>10</sup>:

$$S(j, k) = \frac{1}{W} \left[ \sum_{m=-w}^w \sum_{n=-w}^w [F(j + m, k + n) - M(j + m, k + n)]^2 \right]^{1/2} \quad (1.4)$$

W praktyce, aby wykryć przemieszczenie się jakiegoś obiektu między klatkami filmu, obliczamy odchylenie standardowe dla kilku do kilkudziesięciu klatek filmu dla każdego z pikseli po kolei, to znaczy obliczamy odchylenie standardowe dla wartości pierwszego piksela we wszystkich porównywanych obrazach, drugiego itd. aż do ostatniego piksela we wszystkich analizowanych obrazach.

W wyniku powyższych obliczeń otrzymujemy obraz, na którym obiekt poruszający się lub niewystępujący na wszystkich klatkach analizowanego filmu jest widoczny w postaci jasnych pikseli. Możemy zobaczyć to na przykładzie obrazu z rysunku 4.



Rys. 4. Obraz powstały w wyniku obliczeń odchylenia standardowego dla przykładowego filmu

<sup>10</sup> W. Pratt, *Digital Image Processing*, John Wiley & Sons, Inc., New Jersey 2007.

Obraz z tego rysunku jest wynikiem obliczenia odchylenia standardowego dla 20 klatek filmu, z czego na pierwszych 10 klatkach występuje tylko jeden obiekt – rysunek 5a, natomiast na kolejnych 10 klatkach pojawia się jeszcze jeden obiekt – rysunek 5b. Właśnie ten drugi obiekt, ponieważ pojawia się tylko na 50 proc. klatek filmu, na obrazie powstałym po obliczeniu odchylenia standardowego jest przedstawiony za pomocą jasnych pikseli.



Rys. 5a. Klatka z przykładowego filmu, dla którego obliczono odchylenie standardowe z jednym obiektem



Rys. 5b. Klatka z przykładowego filmu, dla którego obliczono odchylenie standardowe z dwoma obiektami

## Podsumowanie

W rozdziale przedstawiono metody analizy obrazu, które można wykorzystać w systemie wspomagania osób niewidomych w budynku uczelni wyższej.

Metody te dobrano w taki sposób, by optymalnie wykorzystać budowę systemu, nie obciążając jego zasobów, a równocześnie by wyniki pracy systemu były osiągnięte w czasie zbliżonym do rzeczywistego.

W rozdziale omówiono dwa przypadki wykorzystania metod analizy obrazu w takim systemie. Pierwszy przypadek związany jest z rozpoznawaniem tekstu z tablicy. Drugi natomiast dotyczy wykrywania przemieszczenia się przedmiotów między poszczególnymi klatkami filmu.

Z rozdziału wynika, że w obu omówionych przypadkach można zastosować metody analizy obrazu, by wykorzystać je w systemie wspomagającym osoby niewidome w budynku uczelni. Dobór odpowiednich metod analizy pozwoli

na osiągnięcie założonych celów takiego systemu, jakim jest odczytanie tekstu z tablicy bądź wykrycie zmiany położenia obiektu i przekazanie osobie niewidomej informacji w postaci komunikatu głosowego.

## Literatura

1. Gonzalez R.C., Woods R.E., *Digital Image Processing Second Edition*, Prentice-Hall, Inc., New Jersey 2002.
2. Jähne B., *Practical Handbook on Image Processing for Scientific and Technical Applications*, CRC Press, 2004.
3. Jain A., *Fundamentals of Digital Image Processing*, Prentice Hall, New Jersey 1989.
4. Malina W., Smiatacz M., *Cyfrowe przetwarzanie obrazów*, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2008.
5. Young I., Gerbrands J., Vliet L., *Fundamentals of Image Processing*, Delft University of Technology, Delft 1998.
6. Pratt W., *Digital Image Processing*, John Wiley & Sons, Inc., New Jersey 2007.
7. Pratt W., *Digital Image Processing*, John Wiley & Sons, Inc., New York 2001.
8. Ritter G., Wilson J., *Handbook of Computer Vision Algorithms in Image Algebra*, CRC Press, Boca Raton 2001.
9. Stępczyński A., *Metody nawigacji porównawczej*, Gdańskie Towarzystwo Naukowe, Gdynia 2004.
10. Tadeusiewicz R., Korohoda P., *Komputerowa analiza i przetwarzanie obrazów*, Wydawnictwo Fundacji Postępu Telekomunikacji, Kraków 1997.
11. Russ C., *The image processing handbook*, CRC Press, Boca Raton 2007.

## THE USE OF METHODS OF IMAGE ANALYSIS TO ASSIST THE VISUALLY IMPAIRED AT UNIVERSITY

### SUMMARY

This paper will present an overview of image analysis methods from the angle of possibility of use in a system supporting the visually impaired in university buildings. Such a system may provide useful information to the blind. Obtaining this sort of information necessitates, among others, images analysis from the cameras. In

this type of analysis, the key role will be played by the methods selected in a way to provide the lowest energy consumption of the part of the system in which the blind may be equipped. In order to simplify the methods, sensors that provide information might be used. Otherwise the information would have to be obtained by the additional analysis of the image. The result of the review will be to identify the most applicable methods which would be applied in the system.

*Translated by Marek Kannchen*