

## WPŁYW NADMIAROWYCH DANYCH WEJŚCIOWYCH NA JAKOŚĆ MAP ROZBIEŻNOŚCI POZYSKANYCH ZA POMOCĄ UKŁADU PIĘCIU KAMER W KONFIGURACJI EBMCS

Adam L. KACZMAREK

Politechnika Gdańska; Wydział Elektroniki, Telekomunikacji i Informatyki  
tel.: 58 347 13 78 e-mail: adam.l.kaczmarek@eti.pg.gda.pl

**Streszczenie:** Artykuł dotyczy zastosowania układu kamer typu EBMCS (Equal Baseline Multiple Camera Set) składającego się z kamery centralnej oraz kamer bocznych. Na podstawie zdjęć wykonanych za pomocą tego układu można otrzymać mapy rozbieżności (ang. disparity map), które pozwalają na określenie odległości od kamer do obiektów znajdujących się w ich polu widzenia. Mapy te są wyższej jakości niż mapy otrzymane za pomocą kamery stereoskopowej. Artykuł przedstawia zastosowanie EBMCS do otrzymywania map głębi dla wybranych fragmentów uzyskanych zdjęć. Artykuł podejmuje temat wpływu rozmiaru otoczenia badanego fragmentu na jakość wyników. Opisane w artykule eksperymenty zostały przeprowadzone na podstawie dwóch zestawów testowych zawierających zdjęcia roślin.

**Słowa kluczowe:** widzenie stereoskopowe, macierz kamer, mapa rozbieżności, mapa głębi

### 1. WPROWADZENIE

Artykuł porusza temat sposobu obróbki danych w procesie widzenia stereoskopowego polegającym na określaniu odległości w przestrzeni trójwymiarowej na podstawie pary zdjęć wykonanych za pomocą kamery stereoskopowej. Kamera stereoskopowa składa się z dwóch kamer położonych obok siebie i skierowanych w tym samym kierunku. Zdjęcie uzyskane z każdej kamery jest inne, a widoczne obiekty znajdują się na nich w innych miejscach. Ze względu na to, że lokalizacja kamer w kamerze stereoskopowej jest różna, występuje przesunięcie obiektów na zdjęciach względem brzegów zdjęcia. Dzięki widzeniu stereoskopowemu możliwe jest określenie, które obiekty są bliżej, a które dalej od kamery. Wymaga to jednak wykorzystania algorytmów obliczeniowych. Różne algorytmy uzyskują wyniki o różnej dokładności w zależności od ich jakości i jakości danych wejściowych, którymi są zdjęcia z kamer. W artykule przedstawiono oryginalne wyniki badań wpływu nadmiarowego rozmiaru zdjęć na wynik obliczeń odległości przy wykorzystaniu układu pięciu kamer typu EBMCS (Equal Baseline Multiple Camera Set) pełniącego funkcję czterech połączonych ze sobą kamer stereoskopowych [1,2,4].

### 2. WIDZENIE STEREOSKOPOWE

Istnieje wiele czynników mających wpływ na wyniki widzenia stereoskopowego, tj. kalibracja kamer, rodzaj algorytmu użytego do wydobycia danych trójwymiarowych ze zdjęć oraz jakość danych.

Położenie kamer w rzeczywistej kamerze stereoskopowej odbiega od położenia idealnego, w którym osie optyczne kamer są równoległe. W przypadku stosowania pary kamer w układzie kanonicznym, w którym występują dwie takie same kamery umieszczone obok siebie, niedokładności te zmniejszają jakość danych trójwymiarowych otrzymanych na podstawie pary zdjęć. W związku z tym przeprowadza się kalibrację tego rodzaju pary kamer [2,3]. Powszechnie stosowany sposób kalibracji kamer stereoskopowych polega na wykonaniu serii zdjęć znanego wzorca. Na podstawie serii zdjęć określa się zniekształcenia jakie wprowadza układ kamer, dzięki czemu zdjęcia wykonane kamerą stereoskopową mogą zostać skorygowane w celu redukcji zniekształceń.

#### 2.1. Algorytmy widzenia stereoskopowego

Algorytmy widzenia stereoskopowego opierają się na porównywaniu zawartości zdjęć z kamery stereoskopowej. Na zdjęciu z prawej kamery (patrząc z tyłu układu kamer) obiekt widoczny na zdjęciach będzie się znajdował bliżej lewego brzegu zdjęcia niż na zdjęciu z lewej kamery. Różnica w położeniu tego samego obiektu na różnych zdjęciach nazywana jest rozbieżnością. Wartość rozbieżności jest różna dla różnych obiektów w zależności od ich odległości od kamery stereoskopowej. Kalibracja kamer może również wpływać na przesunięcia obrazków. Wartości rozbieżności dla całego zdjęcia pozwalają na utworzenie mapy rozbieżności. W kamerze stereoskopowej jedna z kamer pełni funkcję kamery referencyjnej, a druga kamery bocznej. Poszczególne punkty mapy rozbieżności odpowiadają poszczególnym punktom zdjęcia z kamery referencyjnej.

Na podstawie rozbieżności możliwe jest obliczenie w jednoznaczny sposób odległości obiektów od kamery. W obliczeniach tych korzysta się z dodatkowych informacji takich, jak ogniskowa kamer, dane otrzymane na podstawie kalibracji kamer oraz odległość między kamerami.

Informacje o położeniu obiektów wydobywane są na podstawie identyfikacji wartości rozbieżności dla poszczególnych punktów widocznych na zdjęciu referencyjnym. Informacje te wydobywane są za pomocą algorytmów dopasowujących odpowiadające sobie fragmenty zdjęć (ang. stereo matching algorithms). Jednak algorytmy te nie zawsze zidentyfikują poprawnie ten sam fragment obiektu rzeczywistego na różnych zdjęciach.

Istnieją algorytmy dopasowujące, które analizują dla poszczególnych punktów zdjęcia referencyjnego tzw. okno agregujące, składające się z grupy punktów znajdujących się w otoczeniu punktu analizowanego. Tego rodzaju wycinek zdjęcia jest wyszukiwany na zdjęciu bocznym w obszarze, w którym potencjalnie może występować. Po przeprowadzeniu tego rodzaju dopasowania, nazywanego dopasowywaniem lokalnym, powszechnie stosowane algorytmy przeprowadzają optymalizację globalną. Polega ona na tym, że wartości rozbieżności dla poszczególnych punktów są modyfikowane w zależności od wartości otrzymanych dla innych punktów zdjęcia. Stosowane są również algorytmy identyfikujące na zdjęciach punkty charakterystyczne.

Istnieje ponad 100 różnych algorytmów widzenia stereoskopowego. Tworzone są również ich rankingi. Do rankingów takich należy: Middlebury Stereo Vision Page (<http://vision.middlebury.edu/stereo/>) oraz KITTY Vision Benchmark Suite (<http://www.cvlibs.net/datasets/kitti/>).

## 2.2. Rozmiar zdjęć wejściowych

W związku z przeprowadzaniem przez algorytmy optymalizacji globalnej, wartości rozbieżności dla każdego fragmentu zdjęcia referencyjnego zależą od wyników dla całego obrazka. W wielu zastosowaniach stereoskopii nie ma potrzeby wydobywania map rozbieżności dla całego zdjęcia, ponieważ potrzebne są wyniki dla jedynie pewnego, określonego fragmentu. Nasuwa się więc pytanie, czy bardziej dokładne wyniki dla istotnego fragmentu zdjęcia zostaną uzyskane, jeśli wykona się obliczenia dla całego obrazka, czy też lepiej usunąć części zdjęcia znajdujące się poza fragmentem będącym przedmiotem zainteresowania. Opisujący problem jest przedmiotem niniejszego artykułu.

## 3. UKŁAD KAMER EBMCS

Układ kamer o nazwie EBMCS (Equal Baseline Multiple Camera Set) został opracowany specjalnie do celów wykonywania zdjęć trójwymiarowych składających się ze zdjęcia płaskiego oraz mapy głębi odzwierciedlającej odległość obiektów od kamery. Analizowany w niniejszym artykule układ EBMCS składa się z kamery centralnej oraz czterech kamer bocznych. Odległości pomiędzy kamerą centralną a kamerami bocznymi są takie same. Kamery te znajdują się z czterech różnych stron kamery centralnej. Rysunek 1 przedstawia EBMCS składający się z kamer internetowych MS LifeCam Studio zastosowanych do pozyskania zdjęć użytych w eksperymentach.

Układ EBMCS pełni funkcję czterech połączonych ze sobą kamer stereoskopowych [1,2,4]. Każda z tych kamer składa się z kamery centralnej i jednej z kamer bocznych. Kamera centralna jest wspólna i pełni funkcję kamery referencyjnej. Pozwala to w prosty sposób połączyć dane pozyskane z czterech różnych par kamer. [2]. Dodatkową korzyścią używania układu EBMCS jest to, że odległość pomiędzy kamerami bocznymi a kamerą referencyjną jest taka sama dla każdej kamery bocznej. Dzięki temu dla różnych branych pod uwagę par kamer występować będą takie same wartości rozbieżności dla tych samych obiektów widocznych na zdjęciach.

Korzystając zarówno z EBMCS, jak i pojedynczej kamery stereoskopowej, możemy być zainteresowani pozyskaniem mapy głębi jedynie dla pewnej części obszaru widocznego na zdjęciach. Wycinek ten albo będzie zawierał



Rys. 1. Układ kamer o nazwie EBMCS (Equal Baseline Multiple Camera Set) użyty w eksperymentach

jedynie ten fragment, albo jego rozmiar będzie nadmiarowy i widoczne na nim będą również obiekty położone w otoczeniu badanego fragmentu. W zależności od wielkości wycinka oraz rodzaju algorytmu dopasowywania zdjęć, uzyskuje się różne wyniki, co przedstawiono w rozdziale zatytułowanym eksperymenty.

## 4. EKSPERYMENTY

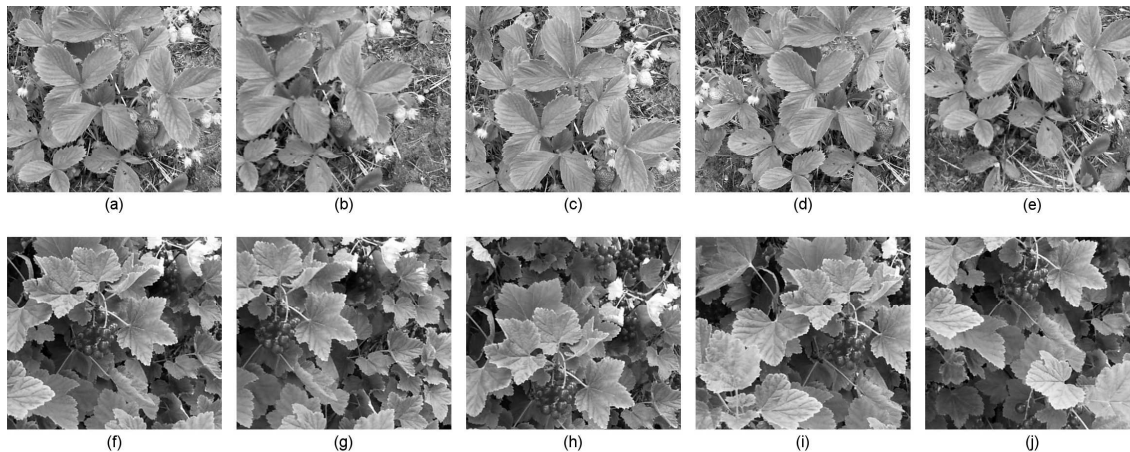
Eksperymenty zostały wykonane na podstawie danych testowych przedstawionych na rysunku 2. Zestawy te użyte zostały również we wcześniejszych eksperymentach z badanym układem kamer. Zawierają dwa zestawy zdjęć roślin wykonanych za pomocą urządzenia przedstawionego na rysunku 1. Zestaw testowy zawierał również prawidłowe wyniki (ang. ground truth), jakie powinien wygenerować algorytm widzenia stereoskopowego. Dane te zostały wydobyte na podstawie ręcznej analizy położenia obiektów na zdjęciach. Dzięki tym danym możliwa jest ocena jakości różnych algorytmów poprzez porównanie ich wyników z wynikami poprawnymi. Wzór 1 określa metrykę BMP (ang. bad matching pixels), na podstawie której przeprowadzona została ocena jakości [2].

$$BMP = \frac{1}{N} \sum_x (|D_M(x) - D_T(x)| > Z) \quad (1)$$

gdzie:  $x$  – współrzędne punktu,  $D_M(x)$  – wartość na mapie rozbieżności w punkcie  $x$ ,  $D_T(x)$  – wartość rzeczywista w punkcie  $x$ ,  $N$  – liczba punktów,  $Z$  – wartość graniczna

W metryce tej występuje wartość graniczna  $Z$ . Jeśli wartość w pewnym punkcie mapy rozbieżności jest nie mniejsza niż wartość poprawna, wówczas uznaje się, że w tym punkcie algorytm podał wynik błędny. Eksperymenty zostały przeprowadzone dla wartości granicznej równej 2.

Korzystając z układu kamer EBMCS możliwe jest zastosowanie różnych algorytmów do wydobywania danych trójwymiarowych na podstawie otrzymanych zdjęć. W tym artykule przedstawione zostaną wyniki stosowania metody EEMM (Exceptions Excluding Merging Method), ponieważ wcześniejsze badania nad układem EBMCS wykazały, że jest to metoda najlepsza [2]. Użycie metody EEMM odbywa się w dwóch krokach. W pierwszym kroku wydobywane są dane trójwymiarowe dla czterech branych pod uwagę kamer stereoskopowych składających się na układ EBMCS. W drugim etapie wyniki z tych czterech kamer stereoskopowych są scalane zgodnie z zasadami EEMM. Algorytm ten opiera się na pomijaniu wartości z tych par



Rys. 2. Zestawy zdjęć użyte w eksperymentach: (a)-(e) Zestaw truskawka, (f)-(j) Zestaw porzeczka; kolejne zdjęcia w poszczególnych zestawach są to zdjęcia z kamery środkowej, prawej, górnej, lewej i dolnej.

kamer, które znacząco odbiegają od wartości pozyskanych z pozostałych par kamer [2].

W pierwszym etapie algorytmu EEMM otrzymuje się mapy rozbieżności z wykorzystaniem algorytmów dopasowujących fragmenty widoczne na parze zdjęć. Na tym etapie możliwe jest użycie różnych algorytmów stosowanych z kamerami stereoskopowymi. W niniejszym artykule przedstawione są wyniki otrzymane dla stosowania algorytmu StereoSGBM (Stereo Semi-Global Block Matching) [5] dostępnego w bibliotece OpenCV [3] oraz algorytmu GC (Graph Cuts using Expansion Moves), którego implementacja udostępniania została na stronie projektu Middlebury Stereo Vision [6]. Opis tego rodzaju algorytmów oraz rozbudowana bibliografia dotycząca tego tematu przedstawiona jest w publikacji [2]. W eksperymentach użyto biblioteki OpenCV w wersji 2.4.13.

Algorytm GC w pełni korzysta z optymalizacji globalnej, algorytm StereoSGBM wykorzystuje taką optymalizację jedynie częściowo. StereoSGBM jest jednym z najczęściej stosowanych algorytmów do stereoskopii z uwagi na to, że biblioteka OpenCV zawiera szybką oraz dobrze przetestowaną implementację. Algorytm GC jest uwzględniany w eksperymentach, ponieważ wcześniejsze badania wskazują, że stosowanie tego algorytmu jest szczególnie korzystne z układem EBMCS [2].

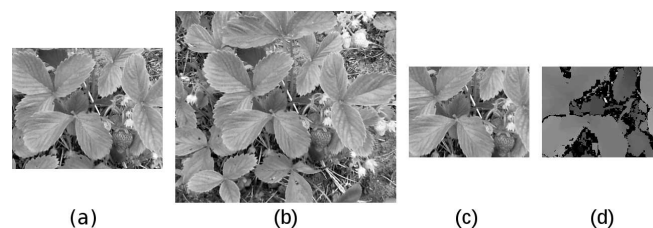
Podczas eksperymentów algorytm EEMM uruchamiany był dla danych wejściowych o różnym rozmiarze. W każdym zestawie danych wejściowych będących fragmentami zdjęć istniał obszar, dla którego obliczana była mapa rozbieżności podlegająca ocenie za pomocą metryki BMP. W przypadku zestawu *truskawka* obszar ten był rozmiaru 240x180 pikseli. Dla zestawu *porzeczka* był to obszar rozmiaru 120x95. Ze zdjęć otrzymanych z kamer wycinane były fragmenty zawierające te obszary oraz obszar dookoła tych fragmentów, który nazywany tu będzie marginesem. Rozmiar ten oznaczany będzie przez  $r$ . Za każdym razem margines zawierał wszystkie punkty oddalone o nie więcej niż  $r$  pikseli od brzegów obszaru poddawanego ocenie.

Badanie miało na celu rozstrzygnięcie, czy wielkość marginesu ma wpływ na optymalizację globalną, w wyniku której otrzymane zostaną różne wyniki dla obszaru będącego przedmiotem analizy. Rysunki 3a oraz 3b przedstawiają przykładowe fragmenty obrazów z centralnej kamery, które utworzone zostały przy zastosowaniu różnej wielkości marginesów. Obszar, dla którego tworzona jest mapa rozbieżności, jest taki sam dla obu tych zdjęć i jest on

przedstawiony na rysunku 3c. Rysunek 3d zawiera przykładową mapę rozbieżności otrzymaną dla przedstawionego fragmentu z wykorzystaniem algorytmu StereoSGBM oraz metody EEMM.

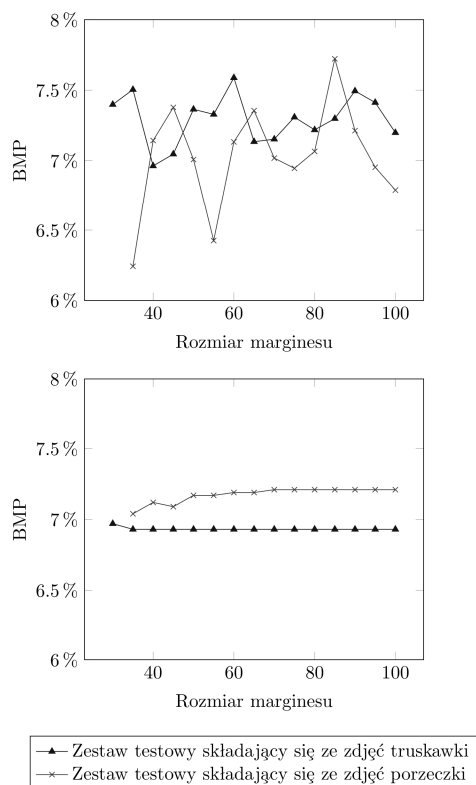
W przypadku zestawu *truskawka* eksperymenty przeprowadzone zostały dla marginesów w zakresie od 30 do 100, co 5 pikseli. Zdjęcia te zawierały obiekty, dla których minimalna rozbieżność wynosiła 28 punktów. Podobnie w przypadku drugiego zestawu minimalna dopuszczalna wielkość marginesu to 35. Eksperymenty przeprowadzone zostały dla marginesów o rozmiarze od 35 do 100 punktów z krokiem 5 pikseli. Wyniki eksperymentów przedstawione są na rysunku 4. Górna część rysunku zawiera wyniki eksperymentu wówczas, gdy razem z metodą EEMM użyty został algorytm GC. Jest to algorytm korzystający z optymalizacji globalnej, który nie zawsze generuje takie same wyniki dla tych samych danych wejściowych. W celu redukcji czynnika losowego wartości podane na wykresie są wartością średnią, obliczoną na podstawie dziesięciu powtórzeń eksperymentu. W przypadku każdego z powtórzeń wyniki odbiegały maksymalnie o 19,4% od przedstawionych wartości średnich.

Eksperyment pokazuje, że istnieją różnice w jakości wyników w zależności od rozmiaru marginesu. Występują zarówno przypadki, w których zwiększenie rozmiaru marginesu spowodowało poprawienie wyników, jak i takie, w których rezultatem było pogorszenie. Średni wynik obliczony dla wszystkich rodzajów marginesu wynosi 7,29% dla zestawu *truskawka* i 7,03% dla zestawu *porzeczka*. Na podstawie eksperymentów można stwierdzić, że wyniki zmieniają się w zakresie powyżej 11% od wartości średniej. W zależności od rozmiaru marginesu i obiektów widocznych na zdjęciach generowane są lepsze lub gorsze rezultaty.



Rys. 3. (a), (b) Przykładowe zdjęcia z kamery centralnej z różnymi rozmiarami marginesów, (c) Obszar dla którego tworzona jest mapa rozbieżności, (d) Przykładowa mapa rozbieżności





Rys. 4. Wartość metryki BMP w zależności od rozmiaru marginesu dla algorytmu GC (górny wykres) oraz dla algorytmu StereoSGBM (dolny wykres)

W ramach eksperymentów otrzymane wyniki zostały porównane z wynikami wygenerowanymi na podstawie pary zdjęć. Użyto zdjęć z kamery środkowej oraz kamery prawej. Eksperymenty przeprowadzone zostały dla takich samych danych wejściowych, jak w przypadku testów pełnego zestawu zdjęć. Dla algorytmu GC średnia wartość metryki BMP wyniosła 14,27 % dla zestawu *truskawka* oraz 8,77% dla zestawu *porzeczka*. Różnice w wartości tej metryki w zależności od wielkości marginesu odbiegały od wartości średniej o ponad 9%. Jest to wartość mniejsza niż w przypadku testów z pełnym zestawem zdjęć, jednak wynik ten został otrzymany dla wartości BMP większych niż w przypadku testów z użyciem pięciu zdjęć. W związku z tym użycie układu EBMCS z algorytmem EEMM poprawia jakość wyników względem wyników z pary kamer, utrzymując na podobnym poziomie zmienność metryki BMP w zależności od wielkości marginesu.

## INFLUENCE OF REDUNDANT INPUT DATA ON THE QUALITY OF DISPARITY MAPS OBTAINED WITH THE USE OF THE FIVE CAMERA SET IN EBMCS CONFIGURATION

The paper describes the usage of a camera array called Equal Baseline Multiple Camera Set (EBMCS) consisting of a central camera and side cameras. The set is designed for taking images for the purpose of acquiring disparity maps which makes it possible to determine distances between a camera set and obtains located within its field of view. These maps have a higher quality than maps acquired using a stereo camera. The paper presents how EBMCS can be used for making disparity maps for only fragments of available images. The paper shows the influence of including in data processing vicinities of the fragment for which disparity maps is generated. Experiments presented in the paper were based on two test data sets containing images of plants.

**Keywords:** stereo vision, camera array, disparity map, depth map.

Innego rodzaju wyniki zostały otrzymane, korzystając z algorytmu StereoSGBM częściowo używającego optymalizacji globalnej. Dla tego algorytmu wpływ wielkości marginesu na wynik jest niewielki. Przedstawione jest to w dolnej części rysunku 4. Podobnie nie zaobserwowano znaczących różnic dla różnego rozmiaru marginesów w przypadku zastosowania algorytmu StereoSGBM do pary zdjęć. Dla zestawu *truskawka* wartość średnia BMP wyniosła 12,84%, natomiast dla zestawu *porzeczka* było to 10,37%.

## 5. PODSUMOWANIE

W związku z wynikami otrzymanymi na podstawie eksperymentów można stwierdzić, że w przypadku badanych algorytmów i użytych w eksperymentach dwóch zestawów danych testowych nie ma potrzeby używania danych wejściowych zawierających szeroki margines, gdy zainteresowani jesteśmy wynikami dla jedynie fragmentu zdjęć wejściowych. Dodatkowo stosowanie większych marginesów zwiększa czas obliczeń, w związku z tym jest korzystne ograniczenie ich wielkości.

## 6. BIBLIOGRAFIA

1. Kaczmarek A. L.: Improving depth maps of plants by using a set of five cameras, *Journal of Electronic Imaging* 24(2), SPIE and IS&T 2015, s. 023018.
2. Kaczmarek A. L.: Stereo vision with equal baseline multiple camera set (EBMCS) for obtaining depth maps of plants, *Computers and Electronics in Agriculture* 135, 2017, s. 23-37.
3. Bradski D.G.R., Kaehler A.: *Learning OpenCV 1st Edition*, O'Reilly Media Inc., 2008.
4. Park J., Inoue S.: Acquisition of sharp depth map from multiple cameras, *Signal Processing: Image Communication* 14 (1-2), 1998, s. 7-19.
5. Hirschmuller H.: Stereo processing by semi-global matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 2008, s. 328-341.
6. Boykov Y., Veksler O., Zabih R.: Fast approximate energy minimization via graph cuts, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23 (11), 2001, s. 1222-1239.