

# Widzenie komputerowe i stereoskopia

---

Rozpoznawanie Obrazów

Krzysztof Krawiec

Instytut Informatyki, Politechnika Poznańska

Na podstawie:

B. Cyganek: *Komputerowe przetwarzanie obrazów trójwymiarowych*,  
EXIT 2002.

# Widzenie komputerowe

Widzenie komputerowe (ang. *computer vision*) – dziedzina zajmująca się interpretacją obrazów trójwymiarowych

- główny problem: mapowanie 3D->2D => utrata informacji (głębi)
- bardziej formalnie: wszystkie punkty leżące na jednej prostej przechodzącej przez ognisko kamery mapowane są na jeden piksel
- cel: uzyskanie mapy głębi (ang. *depth map*) (obraz 2.5 wymiarowy)

# Układ wzrokowy naczelnych

Główne źródła informacji o głębi:

- dysparycja siatkówkowa (ang. *retinal disparity*) ->
- wzajemne przysłanianie obiektów

Te dwa źródła informacji uzupełniają się:

- dysparycja umożliwia ekstrakcję głębi dla części obrazu widocznych dla obu oczu,
- przysłanianie wspomaga segmentację obrazu na obiekty i krawędzie.

Inne źródła informacji:

- ruch gałek ocznych

Co jeszcze wiemy:

- informacja o kolorze ma znikomy wpływ na percepcję głębi

# Dysparycja siatkówkowa – uwagi

- formalnie: dysparycja siatkówkowa jest zerowa tylko dla punktów obrazu rzutowanych na tzw. żółtą *plamkę* („centralny” element siatkówki). Dla innych punktów dysparycja jest zawsze niezerowa.
- w OUN (obszar V1 kory wzrokowej) istnieją neurony zwane *detektorami dysparycji*, reagujące na dysparycję.

Jakie **cechy obrazu** mają zasadniczy wpływ na percepcję dysparycji?

- druga pochodna luminancji (jasności) przecina zero (zero-crossings) [Marr&Poggio79]
- składowe częstotliwościowe sygnału luminancji [Mayhew&Frisby]
- cechy bardziej globalne [Mallot] – wykazano że jesteśmy w stanie skutecznie rekonstruować głębię nawet gdy luminancja zmienia się bardzo powoli.
- najnowsze badania wskazują na charakter hybrydowy, zależny od zawartości obrazu.

Powiązane zagadnienia:

- Stereogramy losowe [Julesz60]

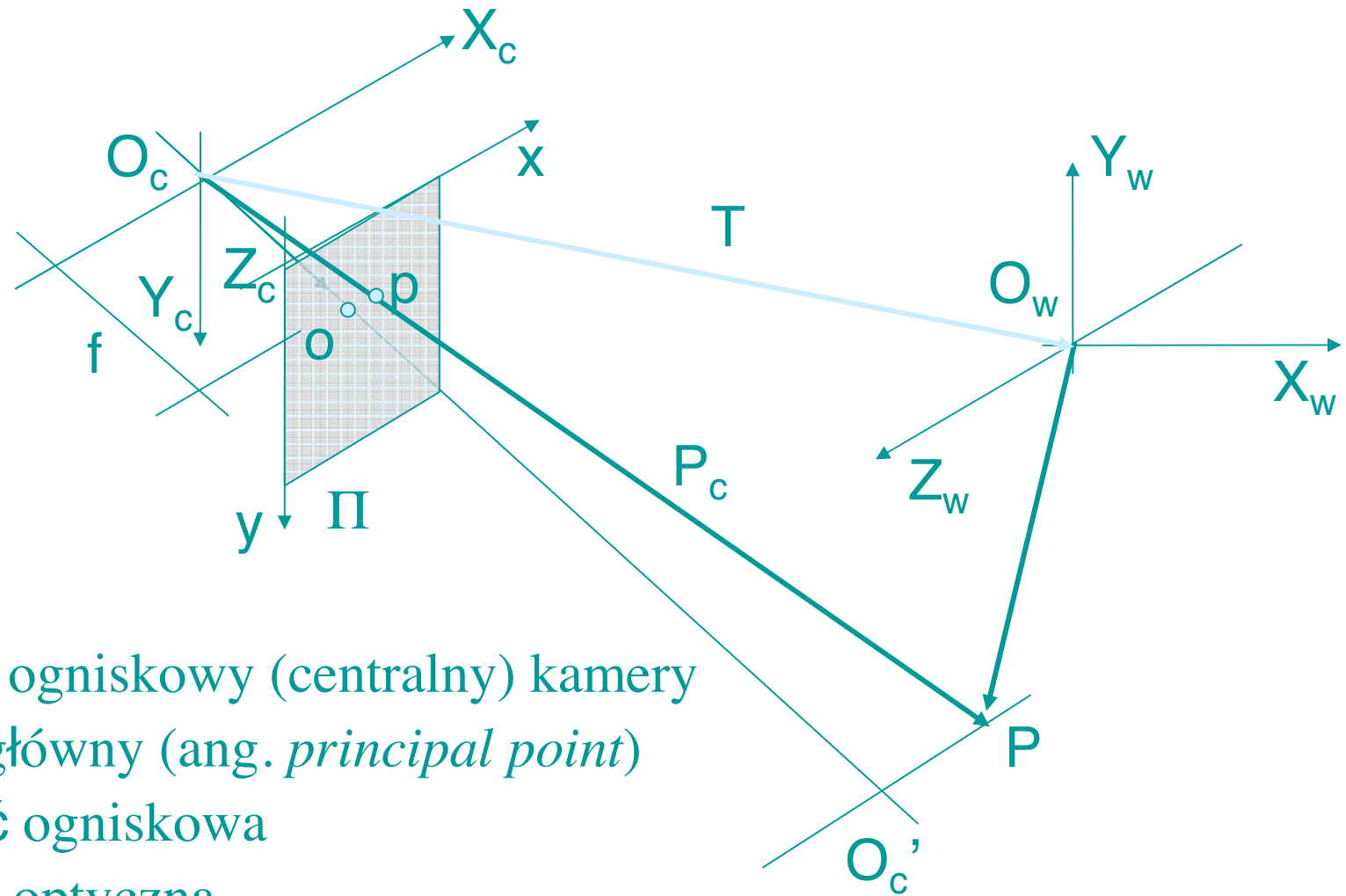
# Stereoskopia

- wykorzystanie dwóch obrazów tej samej sceny (tzw. *stereopara*)
- zalety:
  - Tania
  - Analogiczna do postrzegania człowieka (w obszarze korowym V1 mózgu wykazano istnienie neuronów – detektorów dysparycji, których pobudzenie jest proporcjonalne do niezgodności obrazów pochodzących z lewego i prawego oka)
- wady:
  - wymaga dwóch kamer (zazwyczaj)
  - konieczność kalibracji kamer
  - ograniczona dokładność
  - wysoki koszt obliczeniowy

# Model kamery perspektywicznej

---

# Model kamery perspektywicznej



$O_c$  – punkt ogniskowy (centralny) kamery

$o$  – punkt główny (ang. *principal point*)

$f$  – długość ogniskowa

$O_c O_c'$  – oś optyczna

# Rzutowanie punktu

Dany jest punkt  $P=(X,Y,Z)$  w układzie współrzędnych kamery. Współrzędne jego obrazu  $p=(x,y,z)$  w detektorze kamery można wywieść z podobieństwa trójkątów  $\Delta O_c p o$  i  $\Delta O_c P O'_c$ .

$$x=f \frac{X}{Z}, \quad y=f \frac{Y}{Z}, \quad z=f$$

Wniosek: współrzędna  $Z$  (głębina) wpływa na współrzędne  $x$  i  $y \Rightarrow$  uzasadnienie sensowności stereoskopii.



# Kamera o uproszczonej perspektywie

Gdy założymy że  $X$  i  $Y \ll \text{średnie } Z$  (czyli że rozmiary obiektu są znacząco mniejsze od jego odległości od kamery, co jest zazwyczaj prawdą), wcześniejszy wzór można sprowadzić do:

$$x = \frac{f}{\bar{Z}} X, \quad y = \frac{f}{\bar{Z}} Y, \quad z = f$$

czyli jest to de facto skalowanie przez czynnik  $f/(\bar{Z} \text{ średnie})$ .

[Pokrewny model: model kamery afinicznej]

# Parametry kamery

Dzielią się na:

- zewnętrzne
- wewnętrzne

# Parametry zewnętrzne

- W prostych przypadkach układ współrzędnych kamery i zewnętrzny można utożsamiać. Jednak gdy mamy więcej niż jedną kamerę lub kamera się porusza, nie da się tego zrobić.
- Przejście od zewnętrznego układu współrzędnych do układu współrzędnych kamery: złożenie translacji  $T$  i rotacji  $R$ :

$$P_c = R(P_w - T)$$

- gdzie:

$P_c$  – położenie punktu  $P$  w układzie współrzędnych kamery

$P_w$  – położenie punktu  $P$  w zewnętrznym układzie współrzędnych

$T$  – macierz translacji

$R$  – macierz rotacji (3x3)

$$T = O_w - O_c = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}$$

# Parametry wewnętrzne

- kamera perspektywiczna o obiektywie punktowym ma jeden parametr – ogniskową  $f$ ,
- translacja pomiędzy współrzędnymi kamery a współrzędnymi obrazu: zazwyczaj wybiera się punkt środkowy siatki  $o=(o_x, o_y)$  oraz pewne rozmiary pixeli  $s_x$  i  $s_y$ ; wówczas:

$$x = (x_u - o_x) s_x, \quad y = (y_u - o_y) s_y$$

- zniekształcenia geometryczne wprowadzane przez układ optyczny kamery
  - wynikają z nieliniowości układu optycznego oraz zmienności parametrów w funkcji długości fali:
    - aberracja sferyczna, koma, astygmatyzm, zakrzywienie pola widzenia, zniekształcenia poduszkowe i beczkowe
    - aberracja chromatyczna
- dla porównania: kamera ortograficzna nie ma żadnych parametrów wewnętrznych

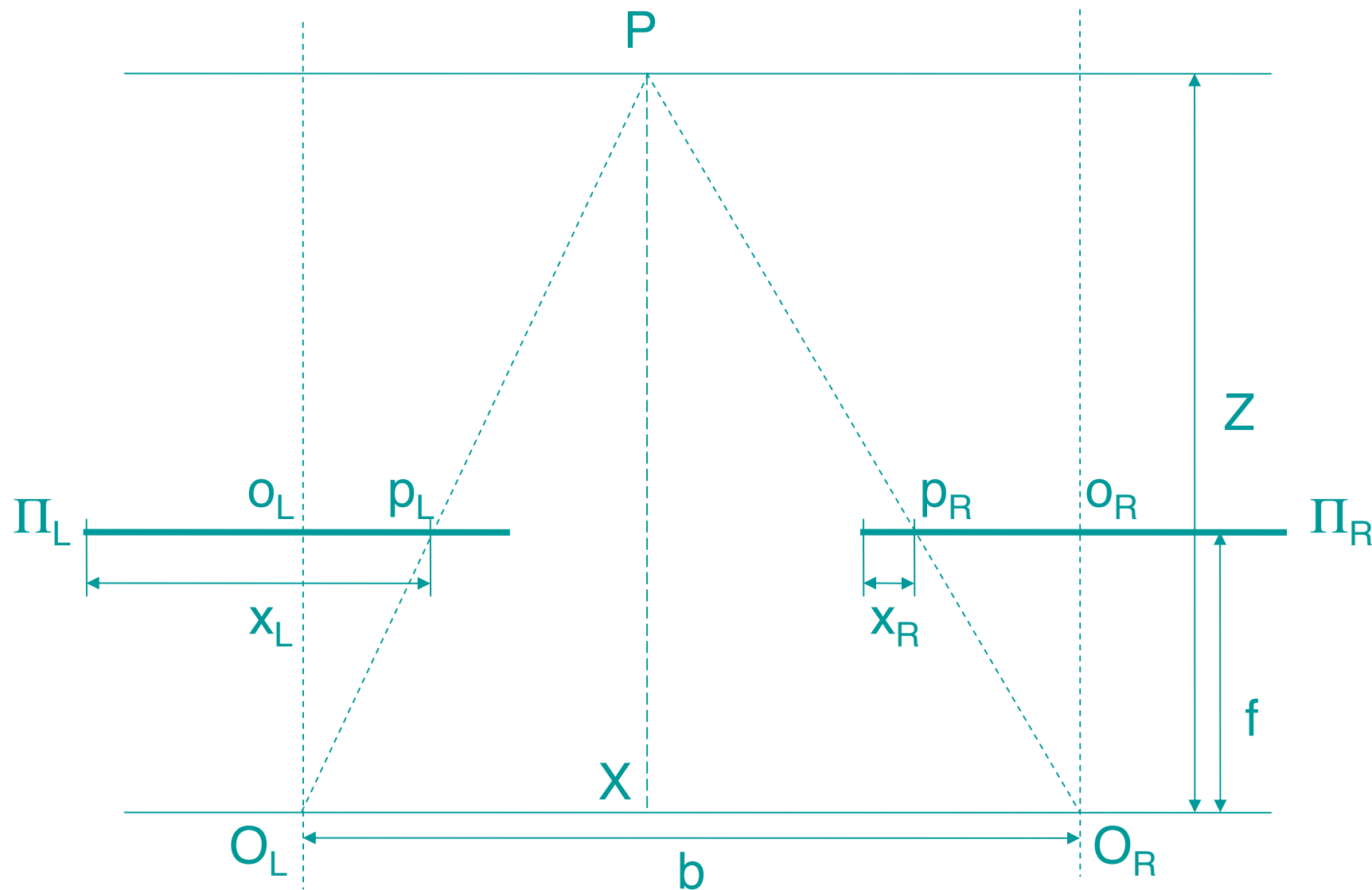
# Kanoniczny układ kamer

Kanoniczny układ kamer to dwie kamery perspektywiczne

- o takich samych parametrach wewnętrznych,
- o równoległych osiach optycznych,
- umieszczone w odległości  $b$  (baza) od siebie.



# Kanoniczny układ kamer



# Definicja dysparycji

Oznaczenia:

- $P$  – punkt w przestrzeni 3D
- $p_L$  i  $p_R$  – jego obrazy w kamerach, odpowiednio lewej i prawej

Korzystając z podobieństwa trójkątów  $\Delta p_L o_L O_L$  oraz  $\Delta PXO_L$ , jak również  $\Delta p_R o_R O_R$  oraz  $\Delta PXO_R$ , dysparycję (horyzontalną) można zdefiniować jako:

$$D_x(p_L, p_R) = x_L - x_R = \frac{bf}{Z}$$

Gdzie  $b$  – odległość między punktami ogniskowymi obu kamer

Można też definiować dysparycję wertykalną  $D_y$ , a następnie zintegrować  $D_x$  i  $D_y$  do dysparycji całkowitej.

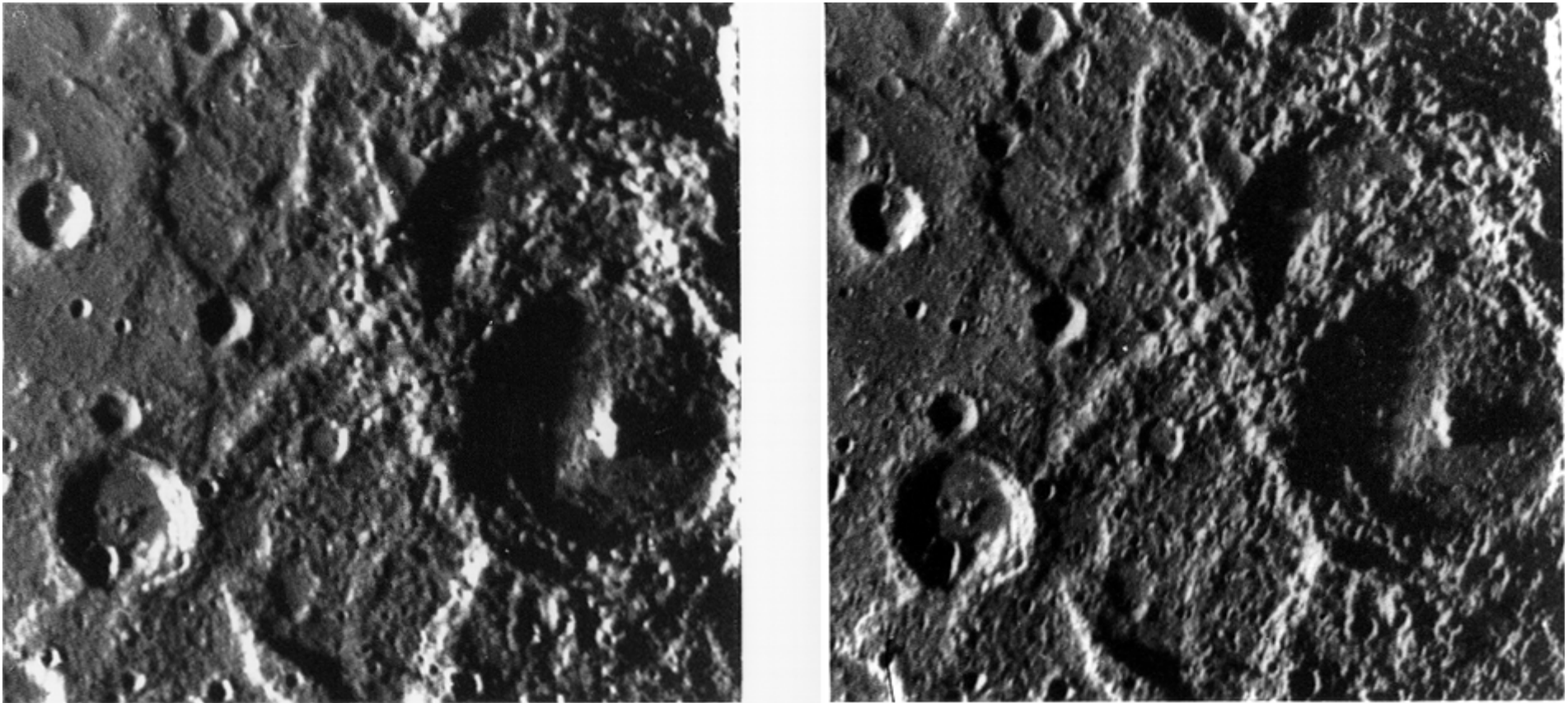


# Przykład 1





## Przykład 2



Mariner 10 stereo pair of Boccacio Crater, Mercury. The right image of this pair was taken during the first flyby, and the left during the second. The crater is 135 km in diameter.



# Metody stereoskopowego przetwarzania obrazów

---

# Zadanie

Mając dane obrazy z lewej i prawej kamery,

- znajdź w nich takie pary punktów które z dużym prawdopodobieństwem są obrazem tego samego punktu sceny,
- przy akceptowalnych nakładach obliczeniowych.



# Metody bezpośrednie

- dopasowujące obszarami (*area-based*), inaczej blokowe lub korelacyjne
  - podstawowa operacja: korelowanie (zgodnie z wybranym kryterium) wybranych obszarów.
- relaksacyjne
  - korelowanie wybranych obszarów + rozwiązywanie problemów niejednoznaczności dopasowań.
  - wprowadza się funkcję „energetyczną”, opisującą „siłę” dopasowania punktów.
- bazujące na programowaniu dynamicznym
  - problem stereowizji sprowadza się do problemu optymalizacyjnego z użyciem pewnej funkcji rozbieżności dopasowań.
  - zakładają zachowanie kolejności punktów.
- gradientowe
  - bazują na obliczeniu pola gradientowego w każdym punkcie obrazu.

# Podstawowy podział

Metody:

- **bezpośrednie** (ang. *direct*) -> :
  - Dysparycja obliczana bezpośrednio z jasności (ewentualnie przefiltrowanej w celu eliminacji szumów).
  - Wynik: *gęsta mapa dysparycji* (estymata dysparycji dla każdego punktu obrazu).
- **metody bazujące na detekcji cech** (ang. *feature based methods*) -> :
  - Podstawą obliczania dysparycji są pewne cechy wyekstrahowane z obrazu (np. linie, narożniki).
  - Wynik: *rzadka mapa dysparycji* (estymata dysparycji tylko w tych miejscach gdzie obecna jest analizowana cecha).
- Typ otrzymanej mapy dysparycji determinuje obszary zastosowań.

# Metody bezpośrednie – c.d.

- neuronowe
  - wykorzystuje się m.in. modele Hopfielda, Hamminga, RBF.
- probabilistyczne
  - budowa statystycznego modelu rozkładu dysparycji w pewnym oknie.
  - model pozwala na obliczenie stopnia niepewności dopasowania.
  - reguła Bayesa pozwala na znalezienie okna minimalizującego tę niepewność.
- dyfuzyjne
  - funkcja energetyczna i dyskretne równanie dyfuzyjne.
  - symulacja pewnego procesu energetycznego.
  - obecność wielu lokalnych minimów funkcji energii -> stosuje się m.in. symulowane wyżarzanie.

# Metody bazujące na cechach

- metoda Marr-Poggio-Grimson
  - detekcja różnego typu krawędzi na różnych skalach szczegółowości, detektorem LoG (Laplacian of Gaussian),
- metoda Shirai
  - dopasowywanie wyłącznie punktów krawędziowych, z pewnym uwzględnieniem ich otoczenia
- metoda tensora strukturalnego
  - polega na detekcji lokalnych struktur w obrazie i dopasowywaniu miejsc charakteryzujących się silną strukturalnością

# Aspekty techniczne algorytmów

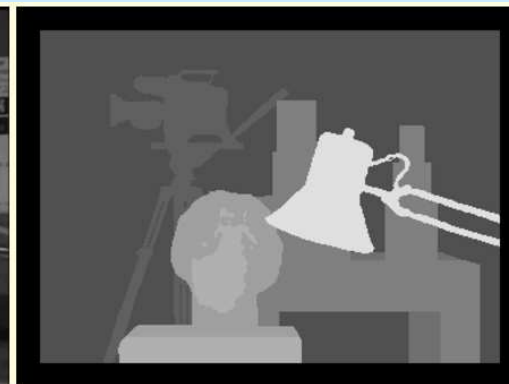
## Algorytmy:

- kroczące po punktach
  - „priorytet dla przestrzeni”: przeglądaj podokna, a wewnątrz każdego podokna doszukuj się różnych dysparycji (zazwyczaj począwszy od najmniejszych)
- kroczące po dysparycjach
  - „priorytet dla dysparycji”: odwrotne zagnieżdżenie pętli: dla różnych oczekiwanych dysparycji przeglądaj całe obrazy





Stereo pair "Tsukuba"



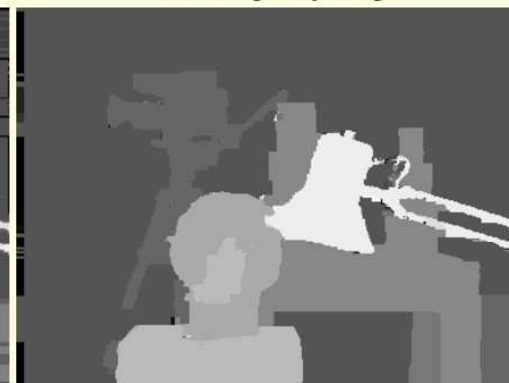
True disparity map



SSD stereo (window 21x21)

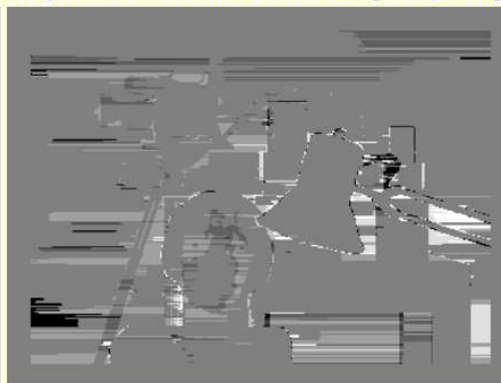
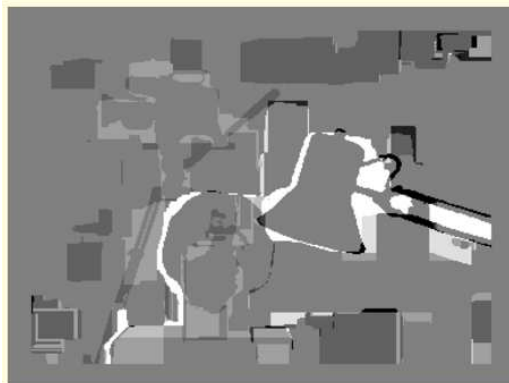


Dynamic programming stereo



Graph-cut stereo

Grey-coded reconstructed disparity maps



Grey-coded signed disparity errors w.r.t. the true disparity map



# Grupy zastosowań

- rekonstrukcja obiektów 3D
- rozpoznawanie obiektów 3D
  - zazwyczaj oparte na modelu -> pomiar podobieństwa mapy głębi i modelu w 3 wymiarach
- kartografia, analiza terenu, w tym zastosowania militarne
  - ciekawy przykład: analiza fal morskich
- systemy nawigacyjne, w tym
  - robotyka
- multimedia
  - synteza obrazów, np.
    - nakładanie obrazów syntetycznych na rzeczywiste
    - generowanie nieznanych widoków obiektów na podstawie widoków znanych

# Stosowanie metod

## 1. W zależności od charakteru obrazów/zastosowania:

- poziom szumów
  - np. wysoki poziom szumów dyskwalifikuje metody bezpośrednie,
- odległość od obrazowanych obiektów i, w konsekwencji, zakres dysparycji
  - np. w obrazach kartograficznych obiekty znajdują się bardzo daleko od kamer[y], dysparycja jest niewielka -> wskazanie na metody bezpośrednie (choć z drugiej strony może wystąpić szum)

# Stosowanie metod – c.d.

## 2. W zależności od stawianych wymagań

- konieczność pracy w czasie rzeczywistym
- wymagana precyzja rekonstrukcji głębi
  - np. często bardzo wysoka w rekonstrukcji obiektów 3D
  - wymaga dokładnej kalibracji kamer

# Metody pomiaru głębi wykorzystujące jeden obraz



# Inne metody pomiaru głębi

Stereoskopia nie jest jedyną metodą pomiaru głębi.

Istnieją metody wykorzystujące jedną kamerę:

- aktywne – wykorzystują osobne źródło promieniowania
- pasywne – wykorzystują oświetlenie naturalne sceny
- inne

# Pomiar głębi - Metody aktywne

- oświetlenie strukturalne (najczęściej regularna krata),
- interferometr (siatka interferencyjna za źródłem promieniowania i przed kamerą),
- laser, radar, sonar – pomiar czasu powrotu sygnału odbitego lub zmiany jego fazy



# Pomiar głębi – Metody aktywne

Zalety: dokładne i często szybkie

Wady: konieczność zastosowania źródła promieniowania:

- kosztowne,
- w przypadku niektórych pasm widma niebezpieczne dla zdrowia
- koszt energetyczny,
- w niektórych zastosowaniach wykluczone:
  - zbyt duża odległość od obserwowanych obiektów,
  - wojskowe, bezpieczeństwo (ang. *homeland security*)



# Pomiar głębi – Metody pasywne

Mogą być wspólnie zatytułowane: „*depth from X*”

- strukturalna analiza cienia (*depth from shading*) – wykorzystanie charakterystyki odbicia/rozpraszania światła do pozyskania informacji o głębi
  - Wady: wymaga silnych założeń, duża złożoność obliczeniowa.
- głębia/kształt z ruchu (*depth from motion*) – wykorzystanie zmian zawartości obrazu przy przemieszczaniu się kamery
  - w pewnym sensie uogólnienie stereoskopii

# Pomiar głębi – Metody pasywne

- głębia z ogniskowania (*depth from focus*) – znajdowanie odległości od obiektu przez zmianę ogniskowej i testowanie ostrości
  - ostrość w lokalnym wycinku obrazu można zmierzyć mierząc udział składowych o wysokich częstotliwościach przestrzennych (nawet np. filtrami krawędziowymi)
  - wady: wymaga kontroli ogniskowej, czasochłonne (nieakceptowalne w wideo)
- głębia z rozogniskowania (ang. *depth from defocus*):
  - korzysta ze spostrzeżenia, że niezogniskowany punkt obrazu ma kształt dysku; dysk ten można opisać pewnym modelem, a pozyskując parametry tego modelu można poznać odległość punktu (pod warunkiem znajomości parametrów optycznych kamery).
  - Matematyczna analiza pokazuje, że wystarczy zaledwie dwa obrazy (o różnym rozogniskowaniu) aby otrzymać mapę głębi

# Metody pasywne

- Zaleta: brak wymagania źródła promieniowania
- Wada: produkują wyniki częściowe, niepełne (np. tzw. rzadka mapa głębi)



# Metody wykorzystujące jeden obraz

*Learning 3-D Scene Structure from a Single Still Image*

Ashutosh Saxena, Min Sun and Andrew Y. Ng

Computer Science Department, Stanford University, Stanford

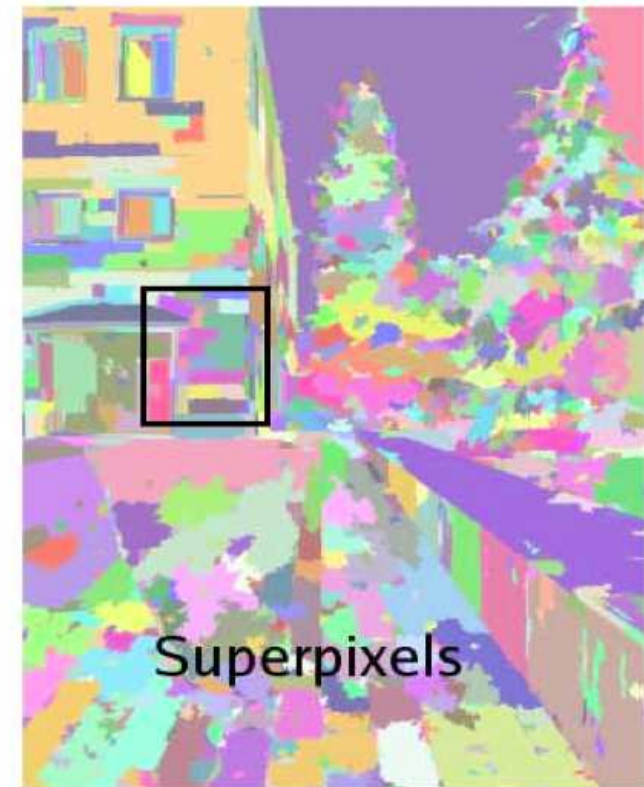
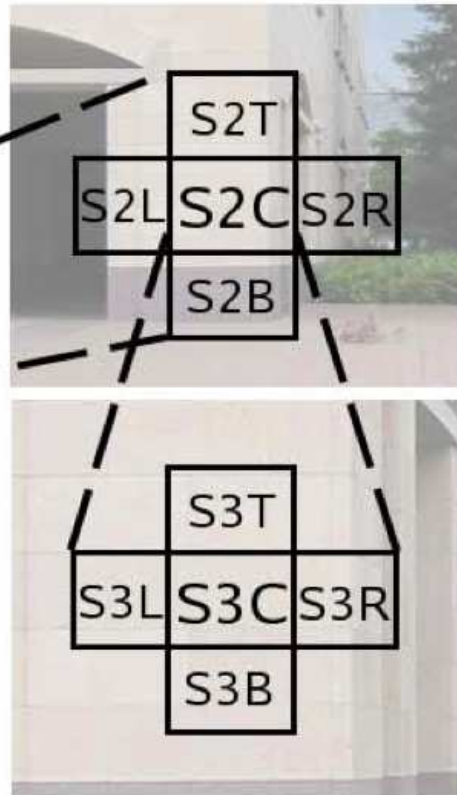
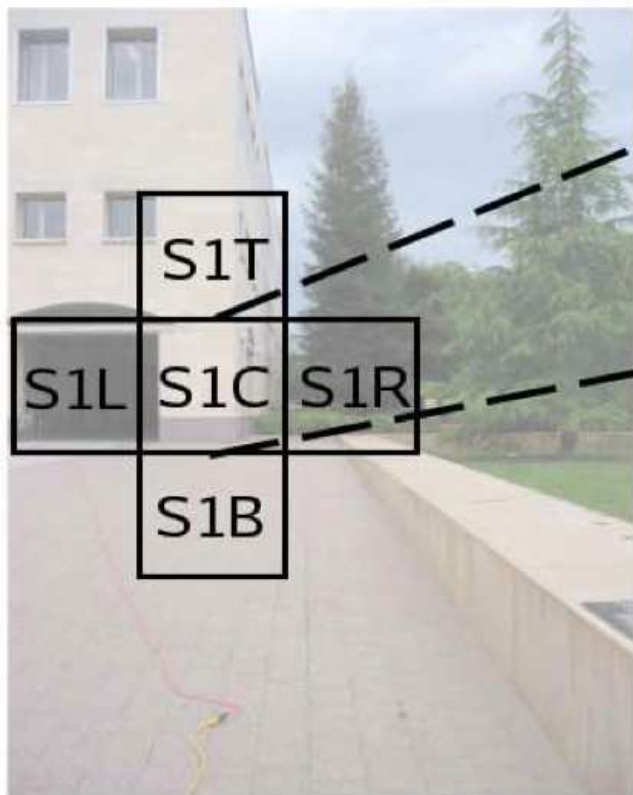
Założenie:

- estymowany model sceny składa się z wielu połączonych ze sobą płaszczyzn



# 1. Wyodrębnianie 'superpikseli'

(oversegmentation)



## 2. Ekstrakcja cech i definicja ograniczeń

Oczekiwana charakterystyka

- Sąsiednie superpiksele prawdopodobnie reprezentują płaszczyzny połączone ze sobą w scenie
- Sąsiednie superpiksele często reprezentują tę samą płaszczyznę
- Proste linie w obrazie 2D reprezentują proste linie w scenie 3D

Dla osiągnięcia dobrych wyników, niezbędne było wzięcie pod uwagę wszystkich powyższych aspektów.



### 3. Uczenie MRF

Model Markov Random Field (przestrzenny proces stochastyczny)

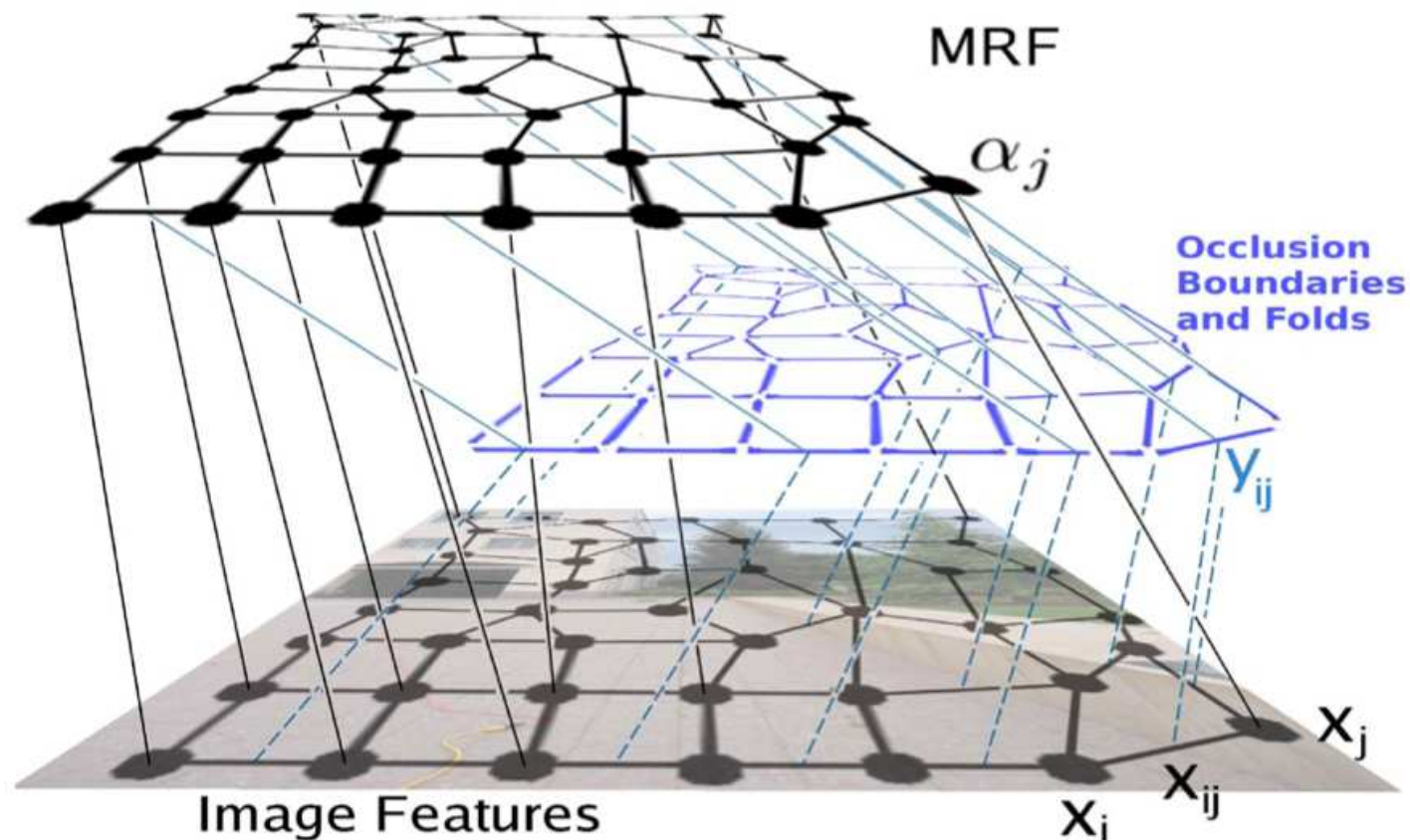


Figure 2. An illustration of the Markov Random Field (MRF) for inferring 3-d structure. (Only a subset of edges and scales shown.)



# Charakterystyka MRF

- Tzw. *plane parameter RF*:
  - zmiennymi są parametry płaszczyzn 3D odpowiadających superpikselom w 2D
- Lokalne cechy obrazu: 524 cechy
  - 4 sąsiednie superpiksele,
  - 3 skale rozdzielczości
- Cechy opisujące relacje pomiędzy superpikselami: otrzymane poprzez dodatkowe segmentacje obrazu
  - 2 skale przestrzenne,
  - 7 zestawów cech.
  - Cecha opisuje czy dana para superpikseli należy do tego samego segmentu w poszczególnych 14 segmentacjach





# Wyniki

Porównanie map głębi pozyskanych z różnych podzbiorów cech

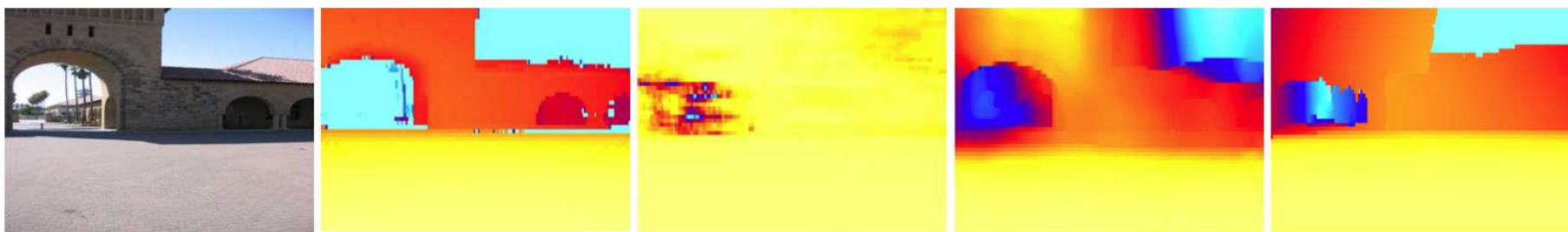


Figure 7. (a) Original Image, (b) Ground truth depthmap, (c) Depth from image features only, (d) Point-wise MRF, (e) Plane parameter MRF. (*Best viewed in Color*)

# Przykładowe wyniki



Original Single Still Image

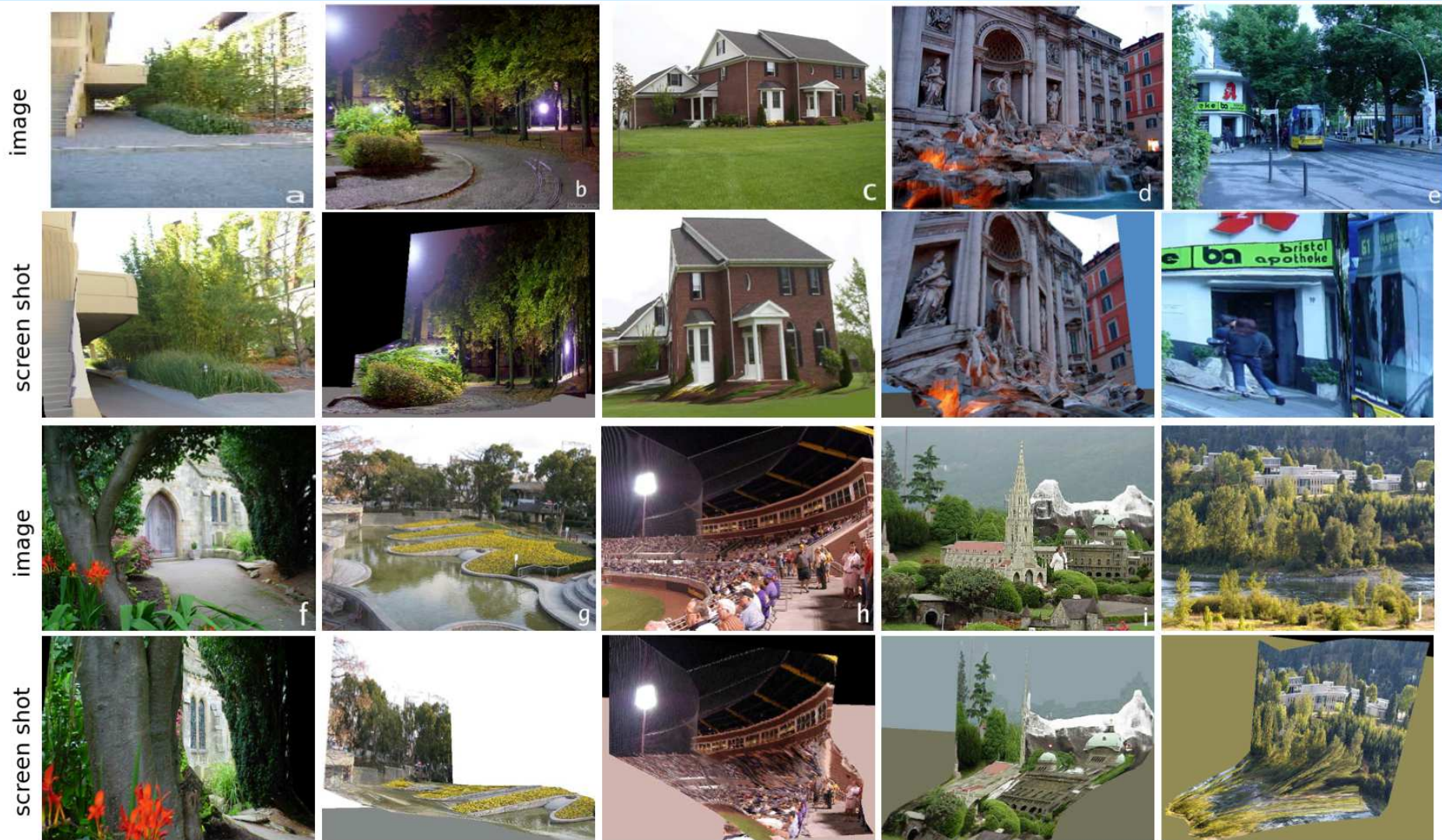


Predicted 3-d model (mesh-view).

<http://make3d.cs.cornell.edu/>



# Przykładowe wyniki (reprojekcja)



The first 7 images (a-g) were evaluated as “correct” and the last 3 (h-j) were evaluated as “incorrect.”





# Nowe specjalizowane urządzenia



# Lytro

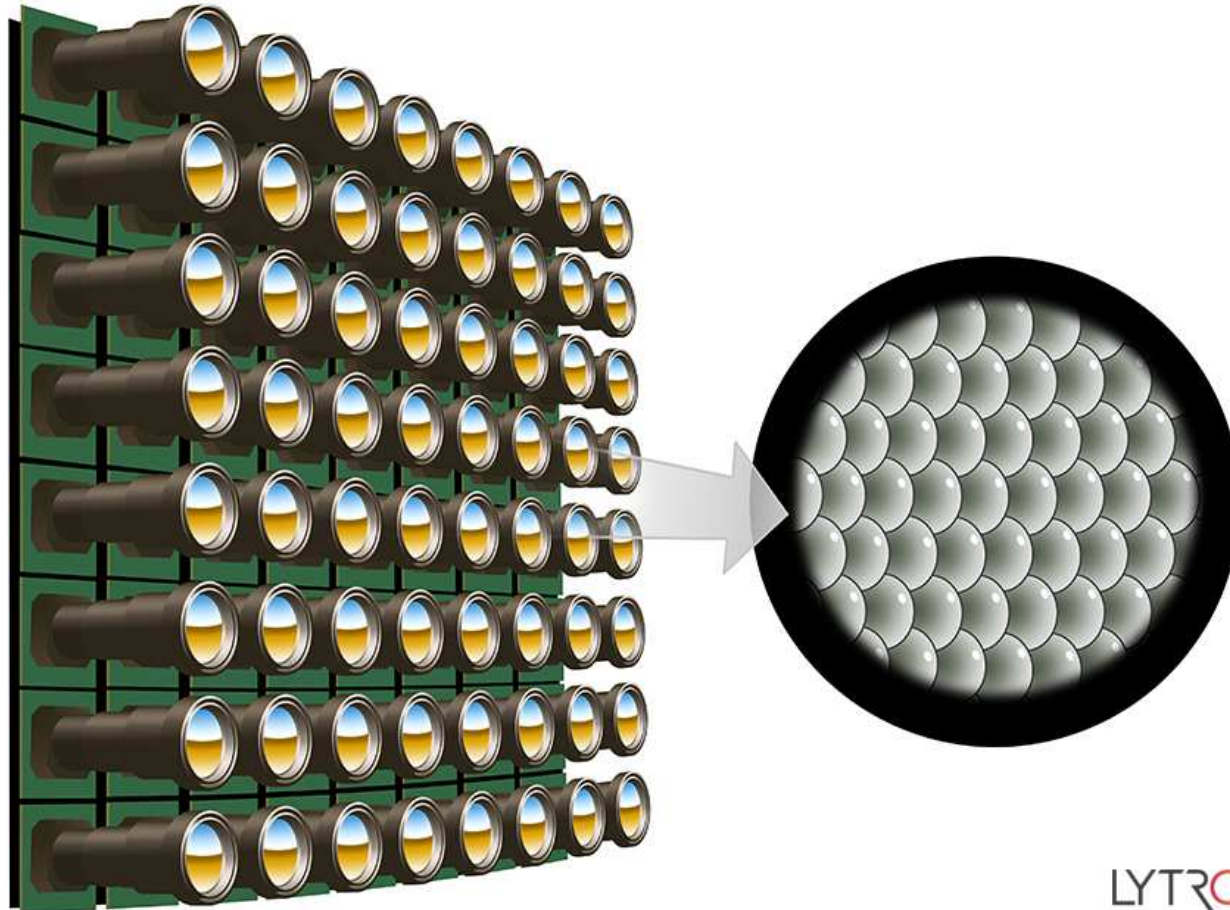
Light field photography (plenoptic photography)

- Zbiera informację nie tylko o intensywności, ale też kierunku światła
- Wykorzystuje matrycę mikrosoczewek (*micro lens array*), które rzutują obraz na matrycę detektorów.
- Rekonstrukcja obrazu odbywa się poprzez transformatę Fouriera.

Cechy:

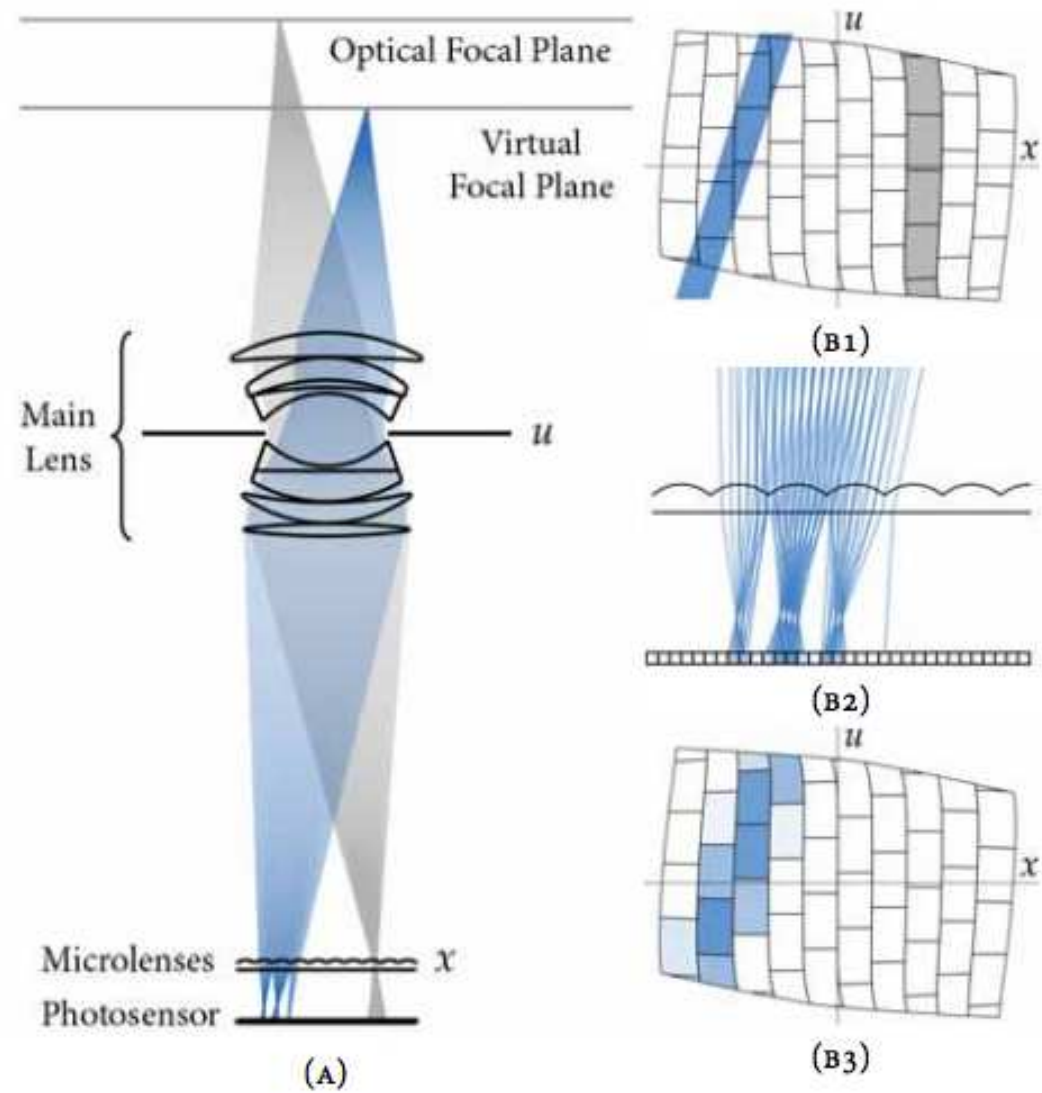
- Możliwość zmiany ustawienia ostrości po wykonaniu zdjęcia.
- Brak konieczności ustawiania ostrości przed wykonaniem zdjęcia, zatem krótszy czas reakcji.
- Możliwość ekstrakcji standardowych obrazów.
- Wysoka czułość (wynikająca z dużej apertury).
- Możliwość rekonstrukcji obrazów 3D.





Multi-camera arrays were the first Light Field capture systems, developed at Stanford University. This massive array has been replicated and miniaturized with the use of microscopic lenslets (microlens array) placed above the sensor in Lytro's ILLUM to capture the Light Field with a single camera. Illustration not to scale.





[case\lytro\CM Movie 2.mov](#)

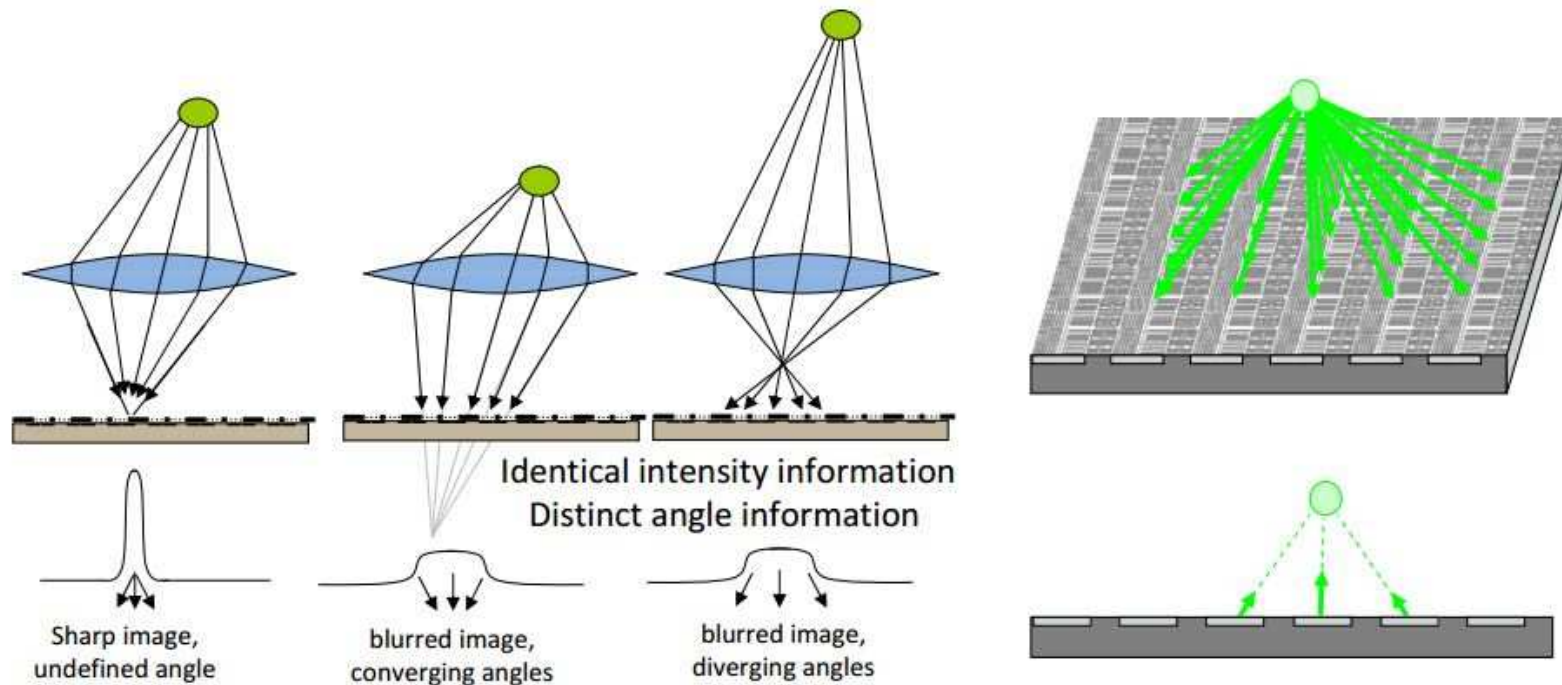
<https://www.lytro.com/>





# Nieco inny pomysł (2012)

## Angle Informs 3-D location



Both with and without a lens, incident angle information can be used to extract the location of a light source in three dimensions. Thus, we want a pixel structure sensitive to ***both the Angle and Intensity of light***