

Percepción de Profundidad: Apareamiento Estereoscópico

Mario Chirinos Colunga, José Ramón Atoche Enseñat.

Instituto Tecnológico de Mérida Av. Tecnológico km. 4.5 Mérida Yucatán, México. www.itmerida.mx
e-mail: chmario@aurea.8m.com, jatoche@labna.itmerida.mx

Resumen:

En este artículo se describe un método de apareamiento estereoscópico para obtener un mapa de disparidad denso y detallado en sub-píxel a partir de un par de imágenes rectificadas. El par de imágenes se comparan por medio de una ventana, cuyo diámetro varía en función de la cantidad de textura dentro de ella. Dicha ventana calcula el costo de apareamiento por medio de la diferencia al cuadrado, almacenando cada costo en un arreglo tridimensional, las oclusiones se determinan mediante un umbral en los costos de apareamiento. Los valores correctos de apareo se mejoran encontrando el valor mínimo de la función de segundo grado descrita por los costos de apareo próximos al mejor valor encontrado, obteniendo así un mapa de disparidad en sub-píxel. Finalmente se corrigen los malos apareamientos mediante un filtro de mediana condicional.

1. Introducción:

Mediante un algoritmo de apareamiento estereoscópico se determina el conjunto de correspondencias correctas entre al menos dos imágenes con el fin de encontrar la profundidad de la escena en base a las diferencias de posición entre un conjunto de correspondencias.

La geometría epipolar limita la búsqueda de correspondencia a solo una dimensión. Mediante la restricción epipolar la correspondencia para un píxel en una imagen se encuentra sobre la línea epipolar de la otra imagen, esta búsqueda se simplifica aun más si se utilizan imágenes rectificadas transformando así la línea epipolar en una línea de exploración horizontal.

La comparación para cada píxel de una imagen se realiza dentro de una ventana mediante alguna función de similitud como la suma de diferencias al cuadrado SSD. En nuestro algoritmo el diámetro de la ventana crece hasta contener la suficiente información para encontrar una buena correspondencia, esto otorga un balance entre detalle y error, el cual es uno de los problemas principales en los mapas de disparidad densos, pues los algoritmos que encuentran un mapa suave, sin errores que generen saltos bruscos de disparidad, tienden a perder detalle y aquellos que obtienen un mapa detallado tienden a ser ruidosos. La búsqueda de correspondencia sobre la línea de exploración se realiza dentro de un rango de disparidad establecido, el mejor valor obtenido por la función de similitud determina a que distancia o disparidad horizontal se encuentra el píxel analizado, respecto a la posición de su correspondiente en la otra imagen. Con cada apareamiento se crea un mapa de disparidad indicando para cada píxel la diferencia de posición a la que se encuentra su correspondencia en la otra imagen. En cada imagen existirán detalles que no se encuentran en la otra pues se encuentran ocluidos: Las oclusiones se determinan mediante un umbral en los costos de oclusión.

El valor de disparidad de cada píxel se mejora encontrando el valor mínimo de la función de segundo grado descrita por los valores de apareo cercanos al valor mínimo, así se obtiene un mapa de disparidad en sub-píxel, donde existe una gama mas amplia de disparidades utilizando fracciones de píxel.

2. Algoritmo

La entrada del algoritmo estereoscópico son dos imágenes rectificadas, derecha e izquierda [Figura 1], las cuales se comparan una a otra para obtener un arreglo tridimensional el cual contiene el costo de apareamiento para cada disparidad, con los valores mínimos de apareamiento se obtienen un mapa de disparidad, en base al cual se encuentran las disparidades en sub-píxel mediante los costos almacenados en el volumen de disparidad, y finalmente los errores en el mapa de disparidad en sub-píxel se eliminan utilizando un filtro de mediana condicional.



Figura 1. Par de imágenes estereoscópicas rectificadas.

2.1 Apareamiento

La diferencia de posición de cada característica en una imagen respecto a la otra es proporcional a su profundidad. Ya que las imágenes de entrada se encuentran rectificadas, esto es son coplanares, el punto correspondiente a (x, y) se encontrara sobre la misma línea y a una distancia o disparidad horizontal d en el punto $(x+d, y)$ en la otra imagen [1], Para encontrar la correspondencia de una imagen respecto a la otra, utilizamos una ventana circular de área Φ centrada en la imagen de interés en (x, y) [2] y otra de igual área en la segunda imagen cuyo centro será desplazado sobre la línea de exploración, línea que comparten las dos imágenes sobre la cual se busca la correspondencia, dentro de un rango de disparidad $(x-D, y)$ a $(x+D, y)$ [Figura 2]. Para determinar cual punto dentro del rango de disparidad es el apareamiento correcto se utiliza la sumatoria de las diferencias al cuadrado SSD [3, 4, 5, 6, 7] como función de similitud para comparar las dos ventanas. Cada píxel de la primera ventana se sustrae de su correspondiente en la otra y la diferencia se eleva al cuadrado, la suma de todos estos valores determina la similitud o costo de apareamiento entre las dos ventanas, entre menor sea el valor SSD mayor será la similitud entre las dos ventanas.



Figura 2. Par de imágenes estereoscópicas son comparadas por medio de dos ventanas, la primera centrada en el punto de interés de la primera imagen y la segunda se desplaza dentro de un rango de disparidad sobre la línea de exploración de la segunda imagen obteniendo el valor de la

comparación para cada valor del rango de disparidad. El diámetro de la ventana varia dependiendo de la cantidad de textura local en ella.

$$SSD = \sum_{(x,y) \in \Phi} (I_1(x,y) - I_2(x+d,y))^2 \quad (1)$$

El área de la ventana Φ , la cual es el número de píxeles dentro de ella, deberá contener la suficiente información para realizar un buen apareamiento [4, 6], si la ventana es pequeña, no contendrá la suficiente información en áreas de baja textura, y si es grande se perderán los detalles. Por lo que el tamaño de la ventana deberá variar en proporción a la cantidad local de textura [Figura 2]. Antes de realizar la comparación entre las dos ventanas se deberá ajustar el tamaño de la ventana.

$$\Phi \propto \sigma = \sqrt{\frac{1}{\Phi} \sum_{(x,y) \in \Phi} (I(x,y) - \bar{m})^2} \quad (2)$$

El radio de la ventana se incrementara hasta que la desviación estándar de las tonalidades en los píxeles dentro de esta sea mayor a un umbral Th_Φ . En la Figura 3 la tonalidad indica el radio de la ventana utilizado para cada píxel, en donde los tonos oscuros corresponden a ventanas pequeñas y los claros a ventanas grandes.

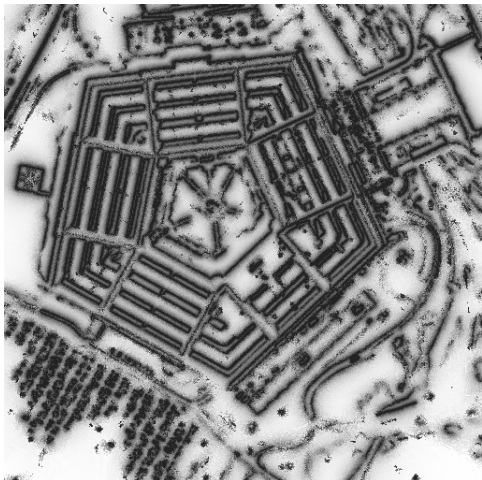


Figura 3. El tono en cada píxel indica el tamaño utilizado para la ventana de comparación (imagen ecualizada).



Figura 4. Mapa de costos de oclusión correspondientes a cada apareamiento utilizado para el mapa de disparidad.

Como el tamaño de cada ventana es diferente para cada píxel, utilizaremos el promedio de las diferencias al cuadrado como función de similitud.

$$Simil = \frac{SSD}{\Phi} = \frac{\sum_{(x,y) \in \Phi} (I_1(x,y) - I_2(x+d,y))^2}{\Phi} \quad (3)$$

Para mejorar el resultado, la contribución de cada píxel a la sumatoria variará de forma gaussiana al alejarse del centro de la ventana.

$$Peso = 2\pi\sigma \cdot e^{-\frac{x^2 + y^2}{2\sigma^2}} \quad (4)$$

De manera que nuestra función de similitud es la siguiente:

$$Simil = \frac{\sum_{(x,y) \in \Phi} (Peso(x,y) \cdot (I_1(x,y) - I_2(x+d,y))^2)}{\sum_{(x,y) \in \Phi} Peso(x,y)} \quad (5)$$

El costo de apareamiento de todos los píxeles para cada valor de disparidad se almacena en un arreglo tridimensional (d, y, x).

$$Costo[x, y, d] \leftarrow Simil(x, y, d) \quad (6)$$

Cada valor del arreglo tridimensional (x, y, d) indica el costo de apareamiento del píxel (x, y) en la imagen de referencia con el píxel (x+d, y) de la segunda imagen. [Figura 5].

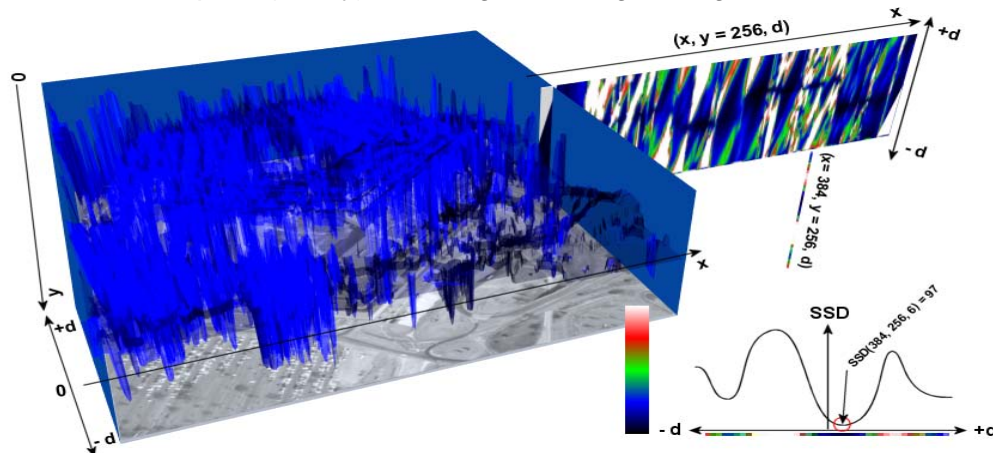


Figura 5. El valor de costo de apareamiento obtenido por la función de similitud para la comparación de la ventana centrada en cada píxel (x, y) de la primera imagen con cada píxel (x+d, y) de la segunda imagen, dentro del rango de disparidad D, es almacenado en un arreglo tridimensional de dimensiones [X, Y, D] en la posición [x, y, d]. Donde X es el ancho de la imagen, Y su altura, D el rango de disparidad y la escala de colores indica el costo de apareamiento

2.2 Mapa de disparidad

Es una imagen donde la intensidad de tono en cada píxel indica a que distancia horizontal se encuentra el píxel correspondiente de una imagen en la otra [1] [Figura 6]. A cada píxel (x, y) en el mapa de disparidad se le asigna el valor de disparidad d, del volumen de costos de apareamiento, para el cual el costo de apareo es mínimo en la coordenada (x, y). De esta manera el mapa de disparidad contendrá el valor de disparidad para cada píxel en la cual se encontró el mejor apareamiento.

En la Figura 6 los niveles de disparidad se indican en escala de grises 256 tonos, la disparidad cero esta indicada por el tono medio, $d_0 = 127$, los valores por debajo de este corresponden a disparidades negativas y las disparidades positivas se indican con tonos mayores a 127.

$$DisMap[x, y] \leftarrow d + 127 \quad (7)$$

Donde d es la disparidad en el punto (x, y) para la cual el costo es mínimo.

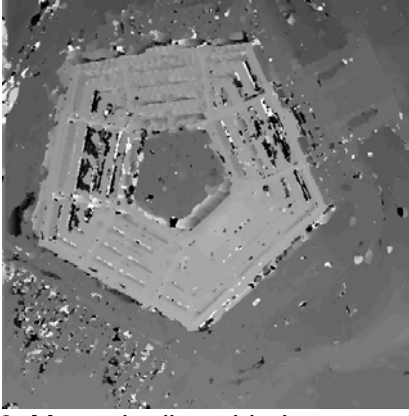


Figura 6. Mapa de disparidad en exactitud de píxel.

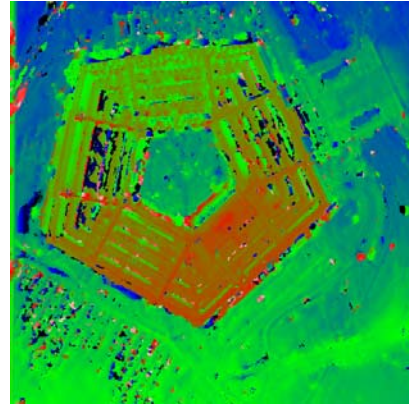


Figura 7. Mapa de disparidad en sub-píxel.

2.3 Mapa de disparidad en sub-píxel

En el mapa de disparidad anterior los niveles de disparidad se encuentran limitados a valores enteros, lo que ocasiona que los cambios en disparidad no sean suaves, se obtenga un mapa tosco y no se aprecien los detalles cuya profundidad esta en disparidades intermedias. Como se observa en la Figura 2 y 5, los valores de los costos de oclusión próximos al valor mínimo describen una parábola, el valor mínimo real para esta parábola puede ser encontrado aproximando los valores a una función de segundo grado por medio de mínimos cuadrados, y hallar los valores del polinomio de segundo grado:

$$Simil = Ad^2 + Bd + C \quad (8)$$

De esta manera se obtienen disparidades fraccionarias (sub-píxel).

Los términos de la ecuación de segundo grado se encuentran resolviendo el siguiente sistema de ecuaciones.

$$\begin{aligned} C \sum_{i=1}^m d_i^0 + B \sum_{i=1}^m d_i^1 + A \sum_{i=1}^m d_i^2 &= \sum_{i=1}^m Simil_i d_i^0 \\ C \sum_{i=1}^m d_i^1 + B \sum_{i=1}^m d_i^2 + A \sum_{i=1}^m d_i^3 &= \sum_{i=1}^m Simil_i d_i^1 \\ C \sum_{i=1}^m d_i^2 + B \sum_{i=1}^m d_i^3 + A \sum_{i=1}^m d_i^4 &= \sum_{i=1}^m Simil_i d_i^2 \end{aligned} \quad (9)$$

El costo mínimo se encuentra igualando la derivada de la función de costos a cero.

$$\frac{\partial Simil}{\partial d} = 0 \Rightarrow 2Ad + B = 0; \quad d_{\min} = -\frac{B}{2A} \quad (10)$$

En el mapa en sub-píxel [Figura 7] se puede observar como se redujeron los apareamientos erróneos y aparecieron detalles que no eran observables en el mapa anterior [Figura 6] pues tienen valores de disparidad fraccionaria.

2.4 Filtro iterativo de mediana condicional.

Si bien el mapa obtenido por el procedimiento anterior contiene los detalles que se encuentran en disparidades fraccionarias, aun contiene malos apareamientos los cuales son áreas pequeñas que presentan saltos bruscos de disparidad.

Eliminamos estos saltos erróneos utilizando un filtro sobre el mapa de disparidad, el cual encuentra la mediana [10] de todos los valores dentro de una mascara centrada en el píxel de interés y reemplaza su valor por el valor mediano si la diferencia entre este y el valor original es mayor a uno.

De esta manera se eliminan los saltos indeseados de disparidad y se conservan los detalles en sub-píxel.

El filtro es aplicado cíclicamente hasta que no encuentre valor alguno que modificar, El mapa de disparidad resultante se muestra en la figura 8.

2.5 Oclusiones

En un par de imágenes estereoscópicas existen detalles que se encuentran en una imagen pero no en la otra, pues se encuentran ocluidos por objetos cercanos debido a la diferencia de perspectiva entre ellas y a los cambios en profundidad. El costo de aparear una característica será muy alto si esta se encuentra ocluida en la otra imagen, pues no existirá similitud entre ninguno de los apareamientos dentro del rango de disparidad D . Para determinar las oclusiones se eliminan todos los puntos del mapa de disparidad cuyo costo de colusión sea mayor a un umbral th_o [5, 10] y se les marca como ocluidos [Figura 9 y 10].

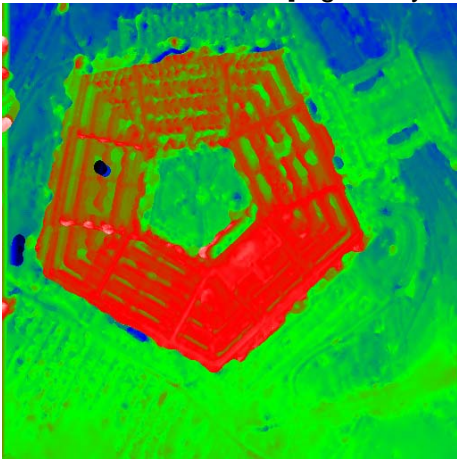


Figura 8: Mapa de disparidad en sub-píxel, obtenido después del filtrado por mediana condicional.

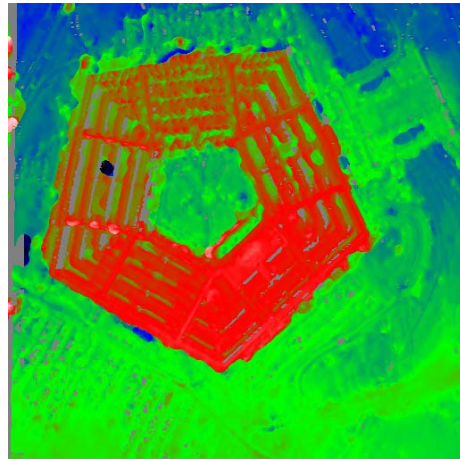


Figura 9 Mapa de disparidad en sub-píxel con oclusiones marcadas en tono gris.

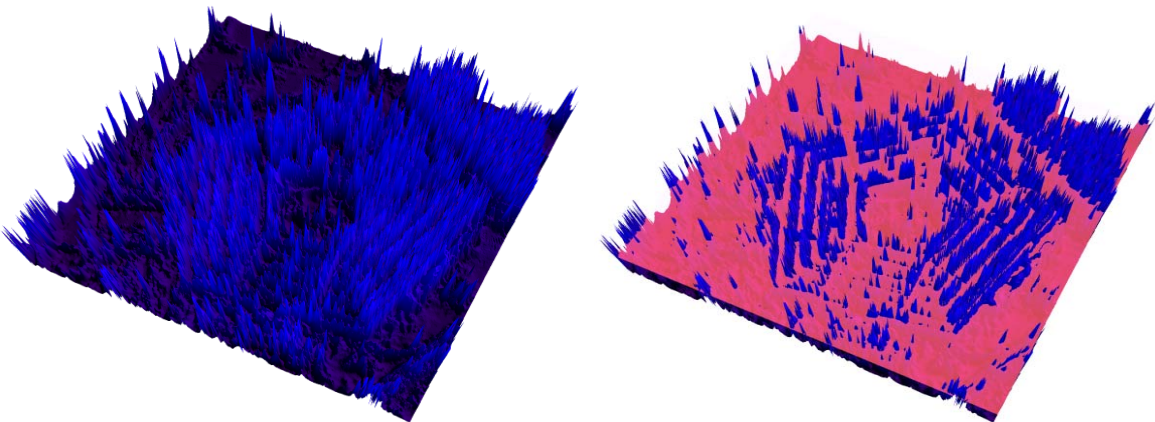


Figura 10. Los costos de apareamiento que son mayores a un umbral th_o se marcan como oclusiones en el mapa de disparidad.

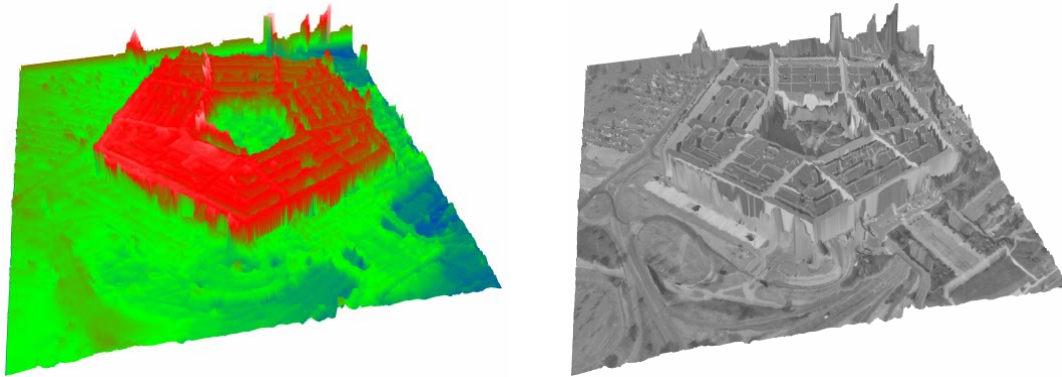


Figura 11: Representación tridimensional del mapa de disparidad.

3. Resultados

Los resultados para la imagen del pentágono, en el primer mapa de disparidad [Figura 6] muestran el detalle obtenido al usar una ventana redonda de tamaño variable en proporción a la cantidad de textura local. Con el Refinamiento en sub-píxel [Figura 7] el detalle se incrementa considerablemente, dejando ver pequeñas variaciones de profundidad que no aparecen en el mapa de disparidad anterior.

Aunque se obtiene un mapa de disparidad detallado existe aun ruido indeseable en la imagen ocasionado por oclusiones y malos apareamientos.

Para eliminarlos se aplica repetidamente el filtro de mediana condicional hasta que se eliminen todos los errores [Figura 8].

Las Oclusiones se determinan por medio de un umbral sobre los costos de oclusión [Figura 10] el cual fue ajustado a 255, los apareamientos con un costo mayor son declarados como ocluidos y se marcan en un tono gris [Figura 9].

Para las pruebas el umbral de textura fue fijado en 20, el de oclusiones en 255, el diámetro menor utilizado para la ventana de comparación en 7 píxeles, el tamaño de la ventana para el filtro de mediana en 5x5 píxeles y la cantidad de valores tomados para el método de mínimos cuadrados al generar el mapa de disparidad en sub-píxel fue de 5 valores de disparidad dos a la izquierda y dos a la derecha de la disparidad de apareamiento. Estos valores fueron calculados experimentalmente.

4. Conclusiones.

Como se puede observar en los mapas de disparidad resultantes, por medio de este tipo de técnicas se puede obtener mapas de disparidad muy detallados y sin ruido.

Si bien la obtención de un mapa de disparidad denso, mediante este método incrementa el tiempo de procesamiento, debido principalmente a que la variación de tamaño en la ventana permite diámetros de tamaño considerable, el tiempo de proceso se puede reducir mediante la paralelización del algoritmo utilizando lenguajes de descripción hardware para implementarlo sobre hardware configurable o utilizando tarjetas aceleradoras graficas GPU las cuales permiten operaciones con matrices de gran tamaño [11, 12].

Referencias.

- [1]. Capitulo. 13, Stereopsis, Paj. 358. Computer Vision: A Modern Approach, by David A. Forsyth,
- [2] Matching Feature Points in Stereo Pairs: A Comparative Study of Some Matching Strategies. Etienne Vincent. School of Information Technology and Engineering, University of Ottawa.
- [3] A Cooperative Algorithm For Stereo, Matching And Occlusion Detection. C. Lawrence Zitnick And Takeo Kanade, Fellow, IEEE
IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 22, No. 7, July 2000
- [4] A stereo matching algorithm with an adaptive window: theory and experiment. Kanade, T. Okutomi, M. Sch. of Computer. Sci., Carnegie Mellon Univ, Pittsburgh, PA. IEEE
- [5] A Maximum Likelihood Stereo Algorithm.

INGEMAR J. COX, SUNITA L. HINGORANI[‡], and SATISH B. RAO.
Computer Vision And Image Understanding, Vol. 63, No. 3, May, pp. 542–567, 1996

- [6] A Fast Area-Based Stereo Matching Algorithm (2002). L. Di Stefano, M. Marchionni, DEIS-ARCES.
- [7] Development of a Video-Rate Stereo Machine.
Takeo Kanade, Hiroshi Kano, Shigeru Kimura, Atsushi Yoshida, Kazuo Oda.
Proceedings of International Robotics and, Systems Conference (IROS'95), Pittsburgh PA.
- [8] Symmetric Sub-Pixel Stereo Matching. Richard Szeliski and Daniel Scharstein.
In Seventh European Conference on Computer Vision (ECCV 2002), Copenhagen, Denmark, May-02.
- [9] A Stereo Vision System for Support of Planetary Surface Exploration.
Maarten Vergauwen, Marc Pollefeys, and Luc Van Gool.
- [10] Calculating Dense Disparity Maps from Color Stereo Images, an Efficient Implementation
Karsten Muhlmann, Dennis Maier, Jurgen Hesser, Reinhard Manner, Lehrstuhl fur Informatik V,
Universitat Mannheim, B6, 23-29, C, D-68131 Mannheim, Germany
- [11] Dense Matrix Algebra on the GPU. Ádám Moravánszky 2004. NovodeX AG.
- [12] Understanding the Efficiency of GPU Algorithms for Matrix-Matrix Multiplication
K. Fatahalian, J. Sugerman, and P. Hanrahan
Stanford University. Graphics Hardware (2004) T. Akenine-Möller, M. McCool (Editors)