Quantization

Floating point   **Integer**

3452.3194   →   3452

32 bit   **8 bit**