A Template Matching Approach for Reference Picture Padding in Video Coding

Nicolas Horst, Priyanka Das, Mathias Wien
Lehrstuhl für Bildverarbeitung RWTH Aachen University

Table of Content

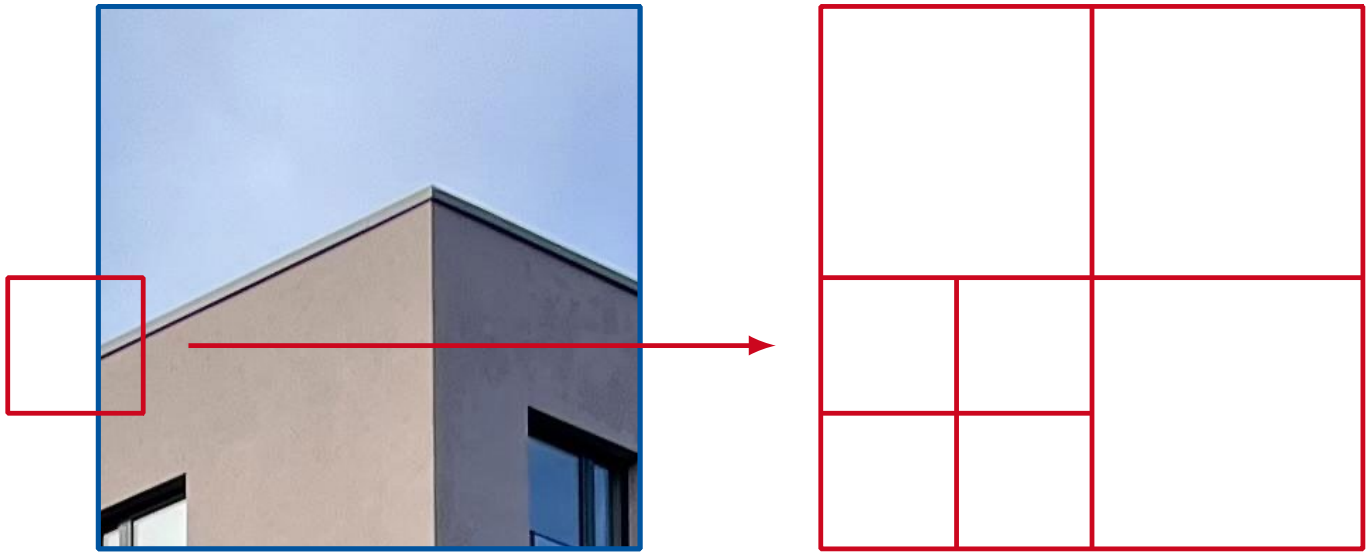
Reference Subpicture Padding

Reference Subpicture Padding

- Subpicture were introduced with VVC
- Independently decodable subpictures are used for viewport adaptive streaming (360 video streaming)
- In this case subpicture boundaries behave like picture boundaries
- Special handling of picture boundaries when inter prediction is used (reference picture padding)

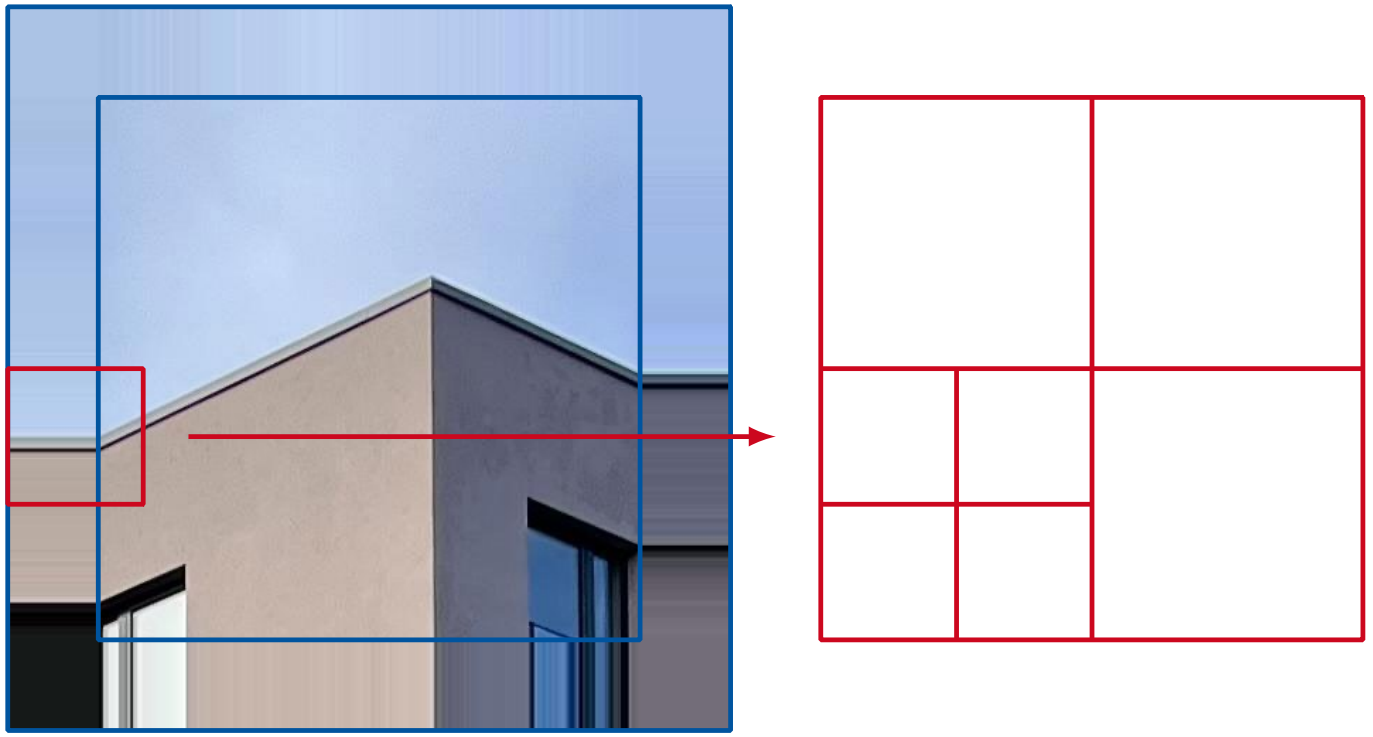
Reference Subpicture Padding

- Subpicture needs to be padded when pixels outside the subpicture are accessed during inter prediction



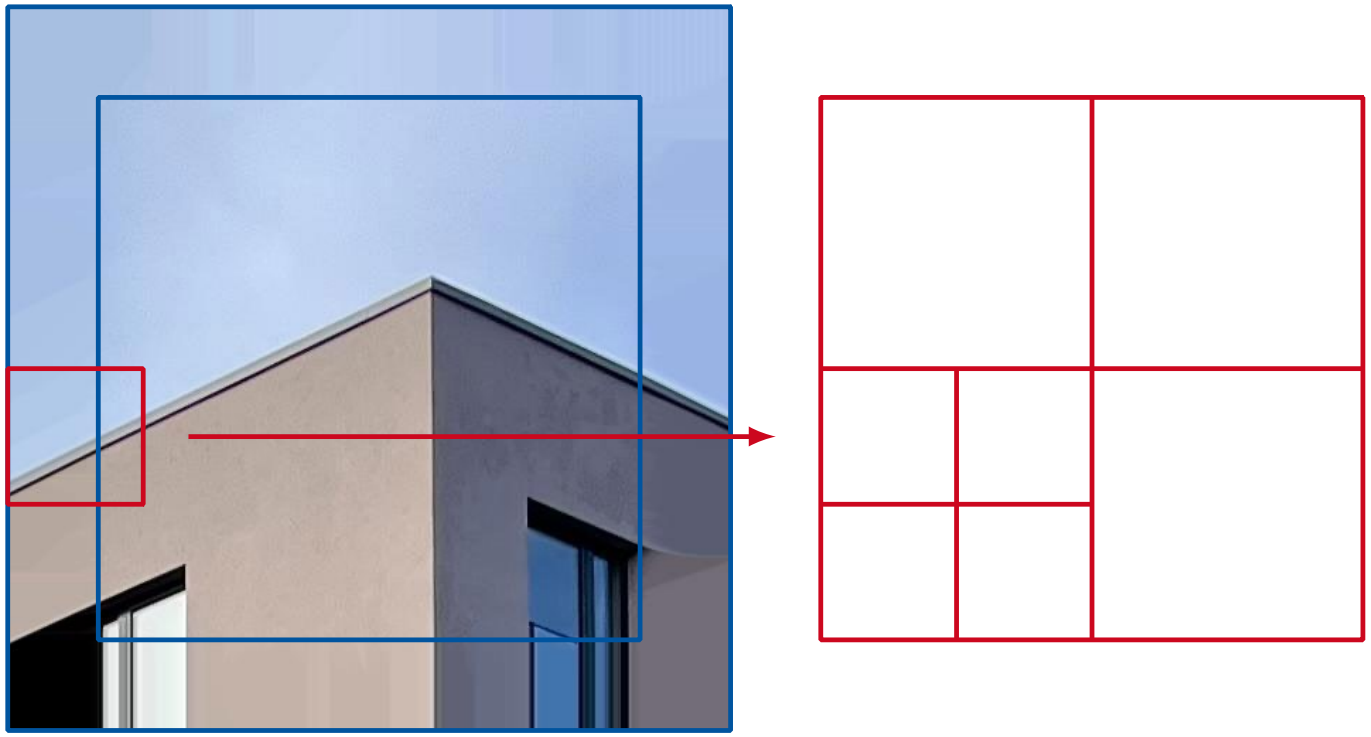
Reference Subpicture Padding

- Subpicture needs to be padded when pixels outside the subpicture are accessed during inter prediction
- In VVC padding is applied by repeating boundary pixels to the outside
- Solution has very low complexity



Reference Subpicture Padding

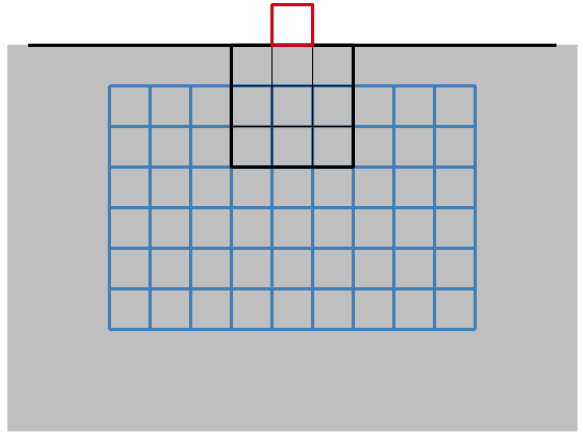
- Subpicture needs to be padded when pixels outside the subpicture are accessed during inter prediction
- In VVC padding is applied by repeating boundary pixels to the outside
- Solution has very low complexity
- Can we do better?



Reference Subpicture Padding with Template Matching

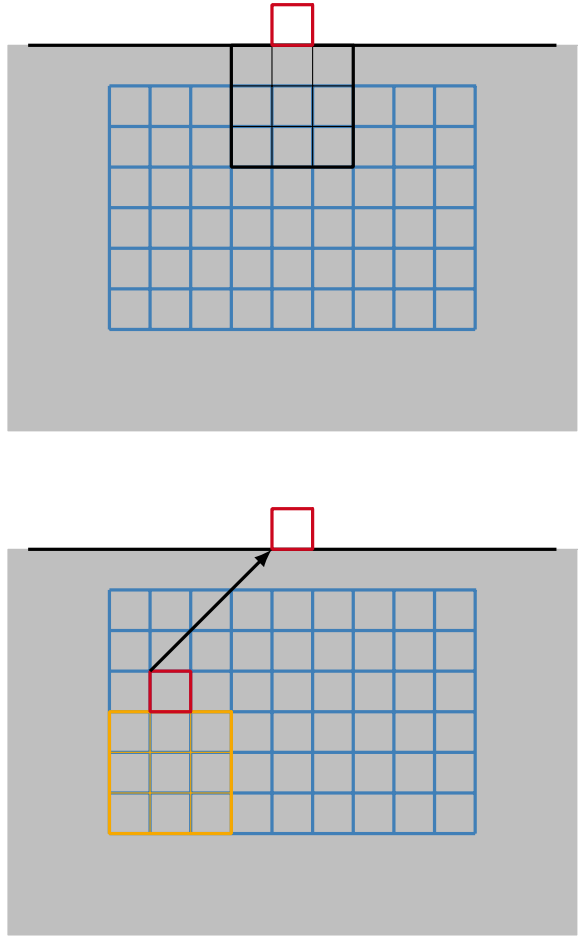
What is template matching?

- For a target block (red border) a template area is defined (black area)
- Search area is defined (blue area)



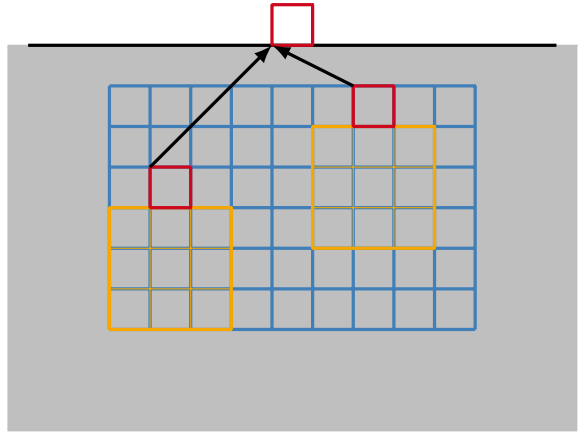
What is template matching?

- For a target block (red border) a template area is defined (black area)
- Search area is defined (blue area)
- Best match(es) for the template area is/are searched in the image (depicted in orange)
- Source blocks (solid red) of the best match(es) are used to predict target block



Averaging multiple candidates

- It has been shown that in the context intra prediction, averaging of multiple candidates can improve results
- Averaging of 5-6 candidates gives the best results for large search area



Template Matching in Video Coding

- Template matching has been used in video coding for prediction
  - Intra prediction:
    - Search area is in the current already decoded part of the picture
    - Averaging of multiple source blocks is beneficial
  - Inter prediction:
    - Search area is in the reference picture
    - Result of search is used to define or refine a motion vector
  - Complexity is a big issue
    - Trading rate for complexity is possible in intra prediction

Template Matching for Reference Subpicture Padding - Algorithm

- Divide the area that needs to be padded into target blocks
- While(area not fully padded):
  - Pad (top, bottom, left, right row):
    - For all target blocks in row:
      - Search for candidates
      - Average best candidates
      - Use as target block

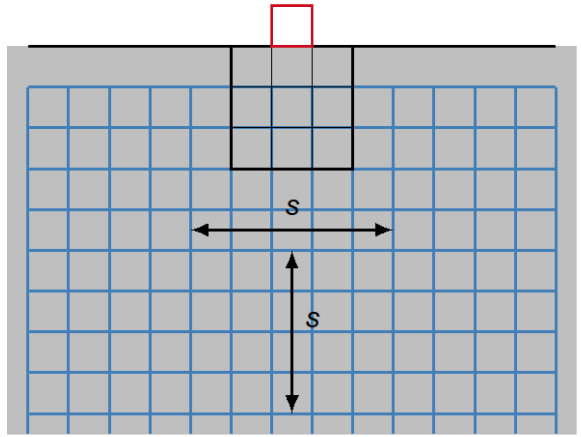
Findings of Pre-experiments

- Pre-experiments:
  - Perform image extrapolation task with algorithm and compare with ground truth
  - MSE is used as metric

- Findings:
  - Decreasing target block size from 2x2, as done in other works, to 1x1 increases quality
  - Increasing template size larger than 3x3 doesn‘t increase quality but complexity
  - Averaging multiple candidates is beneficial (as for TM used for intra prediction)
  - Size of search area has impact on the quality

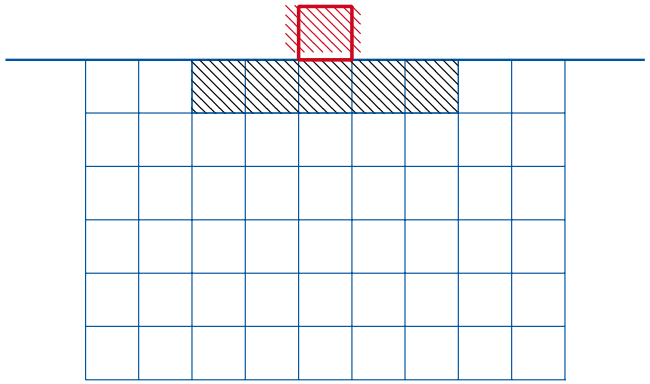
Configuration for the quality setting

- We defined a configuration that focuses on high prediction quality without considering complexity too much
- Search area is $s \times s$ wide and centered around target block at the current border
- We reduce the search area for pixels far away from the original subpicture border to reduce complexity to a feasible level
  - $s = 32,\ \text{distance} \leq 32$
  - $s = 16,\ 32 < \text{distance} \leq 64$
  - $s = 8,\ \text{distance} > 64$
  - 5 best candidates are averaged
  - SAD is used as distortion metric for this and the following setting

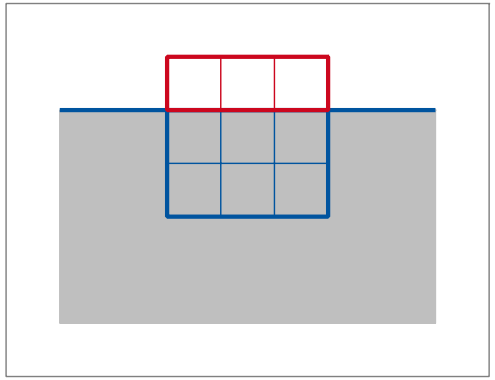


Configuration for fast setting

- Based on the histogram we defined a smaller number of candidates to check (right top figure) to reduce complexity
- We reduced the size of the template in dimension perpendicular to boundary
- We increased the target block in the dimension parallel to boundary, this didn‘t reduce quality in pre-experiments in contrast to the other dimension
- Complexity is reduced by a factor of 330 compared to quality setting with $s = 32$
- Adaptive candidate selection, which selects number of candidates based on threshold that is calculated from minimum SAD of candidates



Source pixel candidates for target pixel



Target block and template

Results

- Simulation on 360 sequences that were converted from ERP to CMP format and partitioned into 384x384 subpictures
- SDR sequences were simulated too
- JVET Common Test Conditions are followed

Visual example
Anchor left
Ours right

360° moving camera set			
sequence	quality	fast	$\text{fast}_{\text{bestC}}$
Balboa	-0.35	-0.20	-0.20
BranCastle2	-0.21	-0.11	-0.09
Broadway	-0.60	-0.36	-0.33
ChairliftRide*	-0.46	-0.29	-0.25
Landing2	-0.34	-0.19	-0.17
SkateboardInLot*	-0.05	-0.10	-0.07
HarborBiking2*	-0.66	-0.41	-0.36
KiteFliteWalking2*	-0.34	-0.20	-0.19
average	-0.38	-0.23	-0.21

* sequences used for pre-experiments

360° static camera set		
sequence	quality	fast
Gaslamp*	-0.05	-0.04
Harbor*	-0.06	-0.04
KiteFlite*	-0.02	-0.01
Trolley*	-0.05	-0.02
average	-0.04	-0.03

* sequences used for pre-experiments

SDR set		
sequence	quality	fast
Campfire	-0.09	-0.02
CatRobot	-0.47	-0.25
DaylightRoad2	-0.48	-0.25
FoodMarket4	-0.64	-0.45
ParkRunning3	-0.35	-0.09
Tango2	-0.85	-0.59
average	-0.48	-0.28

Conclusion

Conclusion

- Template Matching can be successfully used for reference subpicture padding to improve coding performance
- Complexity is an issue with template matching
- Preliminary results show that complexity can be reduced by ~330 times while maintaining about 60% of the BD-rate reduction, subject of current research to improve this
- Artifacts at subpicture boundaries can be reduced

Thanks for your attention

Adaptive candidate selection

- We found that averaging over multiple candidates reduces quality for the fast setting
- We assume that this is due to the low number of candidates for the fast setting

- We introduced a candidate selection process that works as follows:
  - Multiply the minimum SAD with a factor α
  - Average all candidates with corresponding distortion d for which $d < \alpha \cdot d_{\min}$

- Used in the fast setting, $\text{fast}_{\text{bestC}}$ uses only the best candidate

Candidate Selection

- We computed histograms of the relative position of candidates in pre-experiments
- Example histogram for KiteFliteWalking2 sequence is given on the right, logarithmic scale
- For all tested sequences, a T-shaped structure can be observed