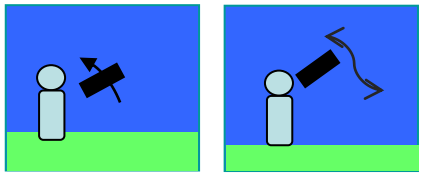
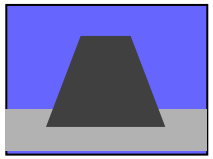
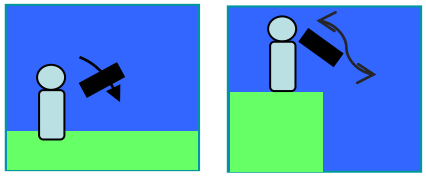
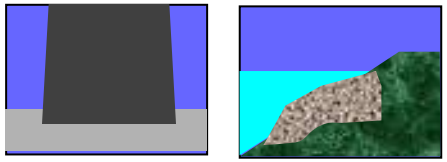
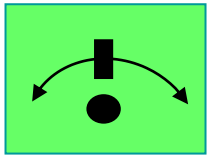
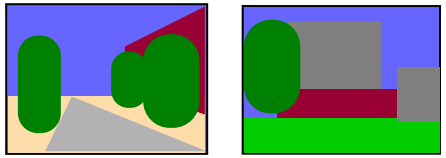
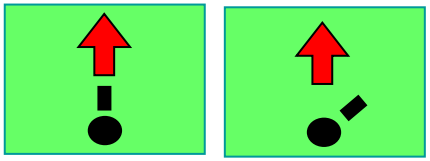
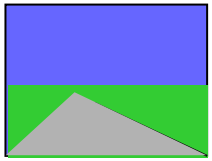
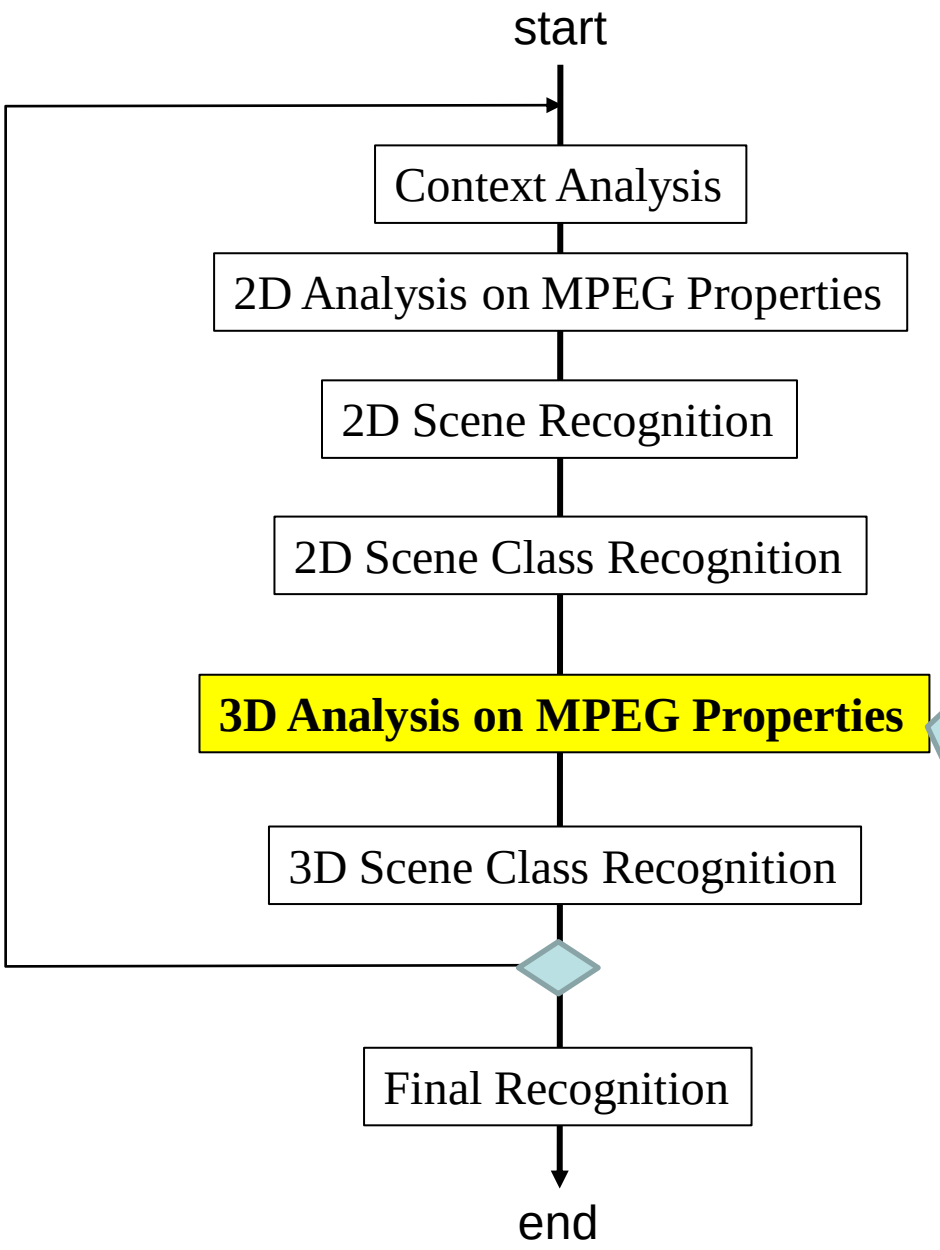
An Investigation on MPEG-based Scene Class Recognition using Factorization

Mikio SASAKI†
E-mail: †{czr04566@nifty.ne.jp}

Primary Camera Motions for the Landscape Scenes

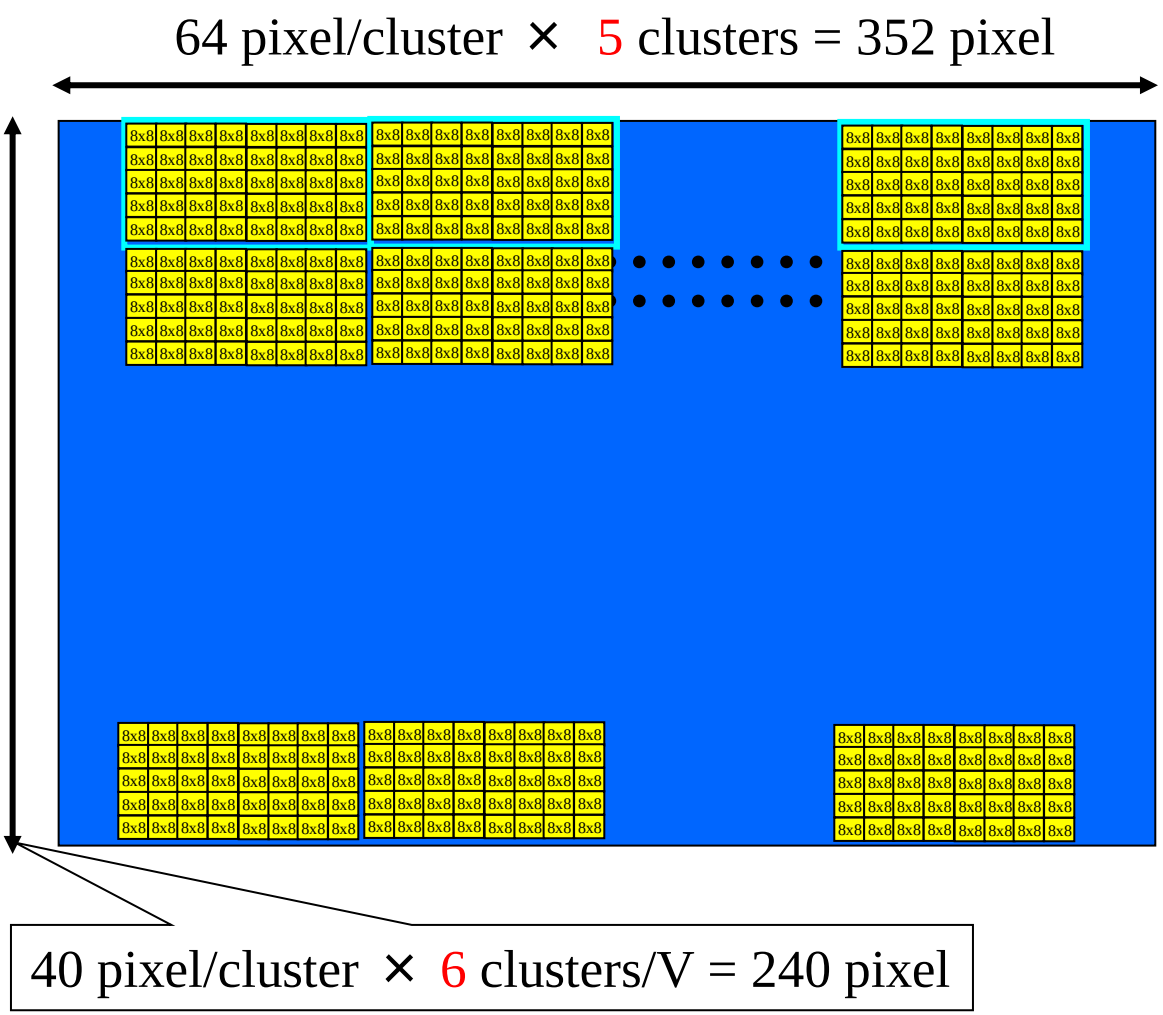
Type	Motion Model	View Model	View Example
UP			 
DOWN			 
AROUND (PAN)			 
THRO-UGH			 

Overview and Problem Definition



- Rank problem occurs when using factorization for the short sequence
- Feature points tracking in pixel resolution is rather heavy for the scene class recognition task

Block-based Factorization (Type-1)



30 clusters in a plane

MERITS: Fast, Low Rank and Low Delay

Extension of the Notion of Cluster for the Block Based Factorization

The cluster is not restricted to the block shaped type. Only the way of decomposition of the initial image plane into the group of blocks specify the definition of each cluster. This is performed under the condition that the number of blocks in each cluster (namely, the number of feature points in the cluster) cannot exceed NK, where, N means the number of the observed attributes and K means the number of observation frames. This is because the time series of each cluster forms the corresponding measurement matrix which row size is NK (in this case, N=2). The tracking operation of each cluster is performed by using motion vectors.

Moreover, in general, the "decomposition" can be replaced with just the "mapping". Therefore, there are no spatio-temporal restrictions to select each block for the corresponding cluster. That is, the duplicated mapping to the different clusters is also possible for each block of pixels. This operation might be effective for the total integration of each partially reconstructed 3D information.

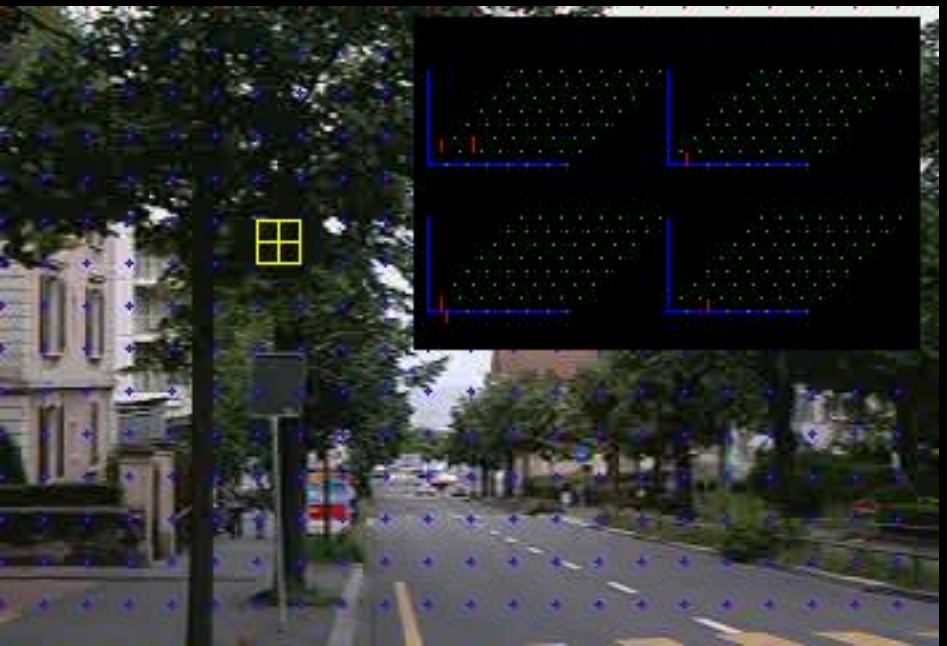
From the computational viewpoints, the administration of "appear and disappear" from frame to frame regarding each feature point is rather complicated problem, even if the observation time is supposed to be short.

In view of more extensions, it is fascinating to consider the case when N is greater than 2. It becomes the generalized or extended factorization which is expected to handle multi-attribute measurement matrix. The corresponding experiments are under way. The automatic selection of the target dimension from the output matrix would be the most critical problem.

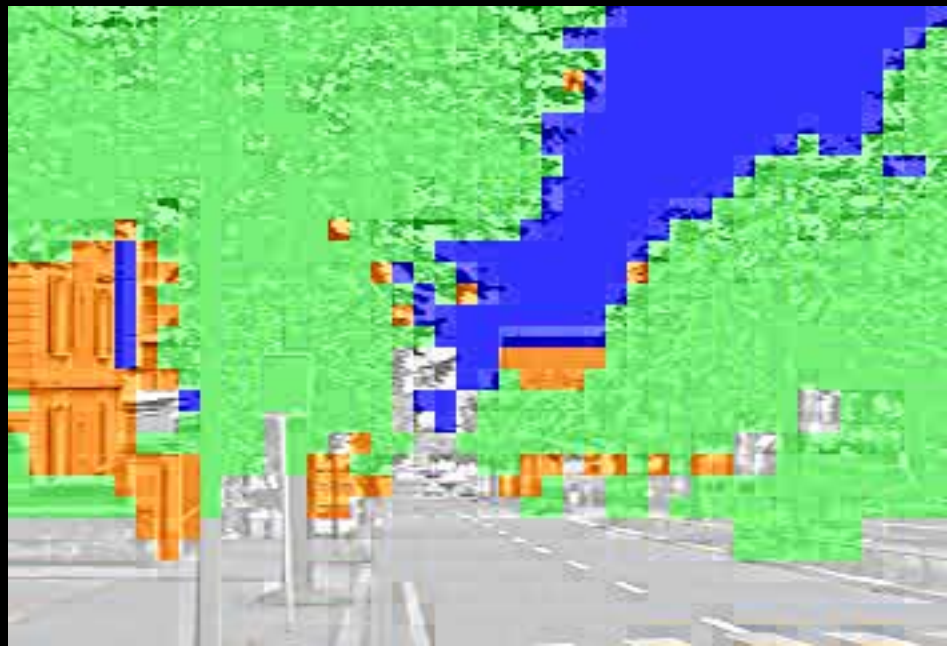
Original

High Power Blocks for Frame Difference

Motion Vectors and DCT coefficients



2D Recognized Results

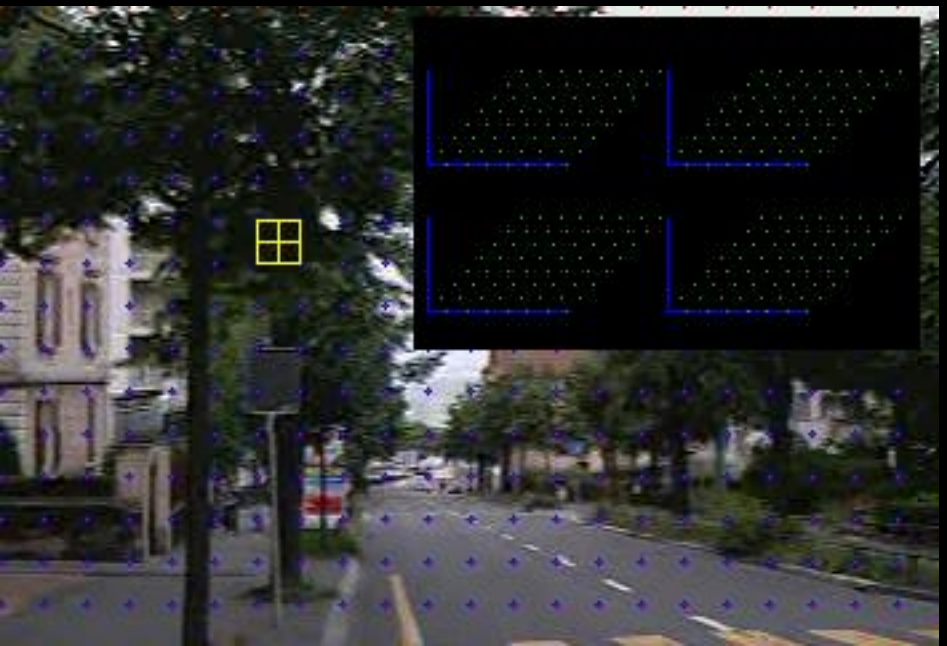


Original

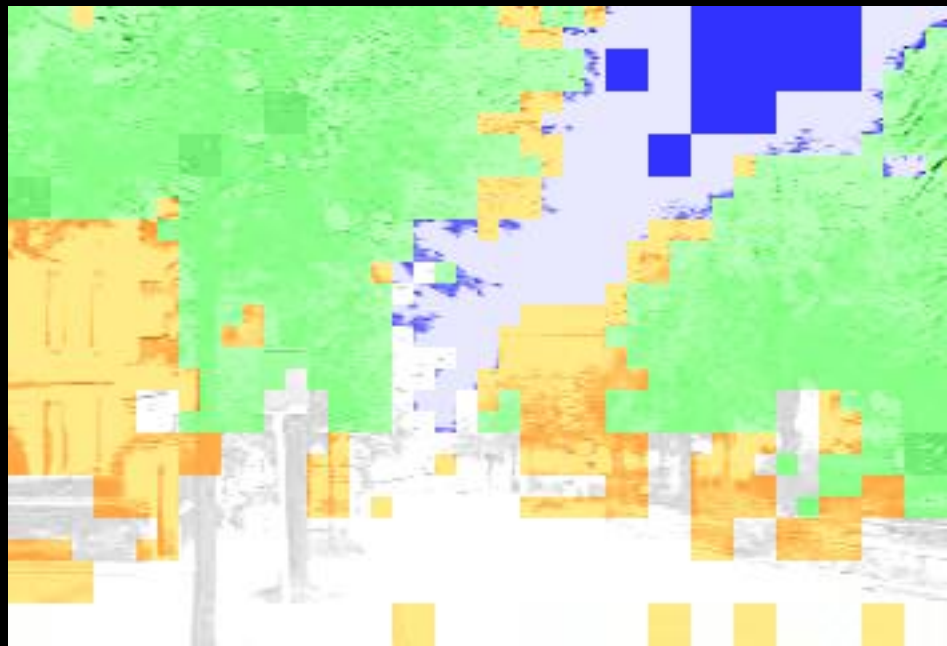
16P6

High Power Blocks for Frame Difference

Motion Vectors and DCT coefficients



2D Recognized Results

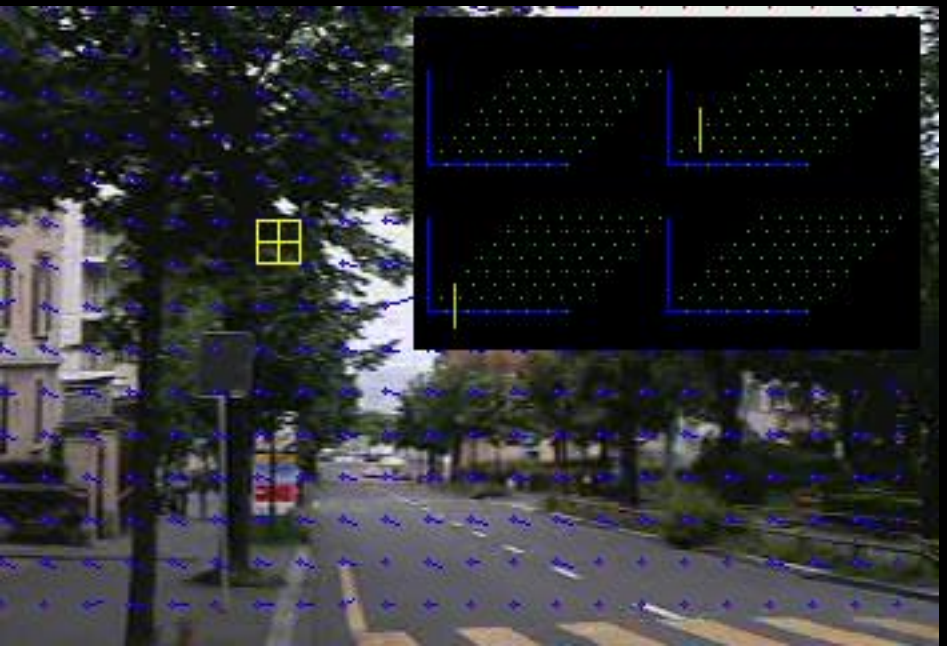


Original

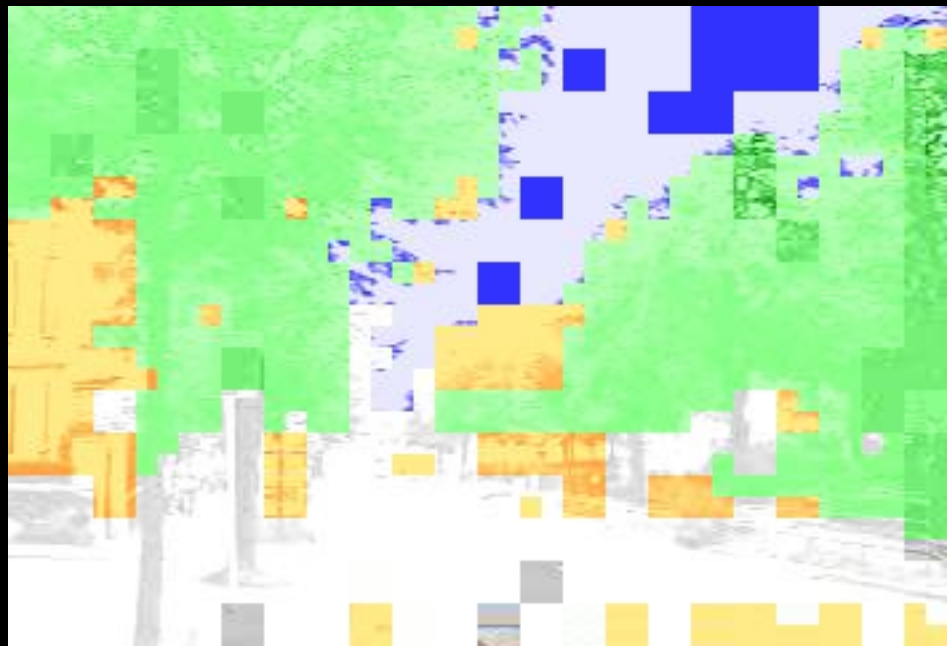
28P10

High Power Blocks for Frame Difference

Motion Vectors and DCT coefficients



2D Recognized Results



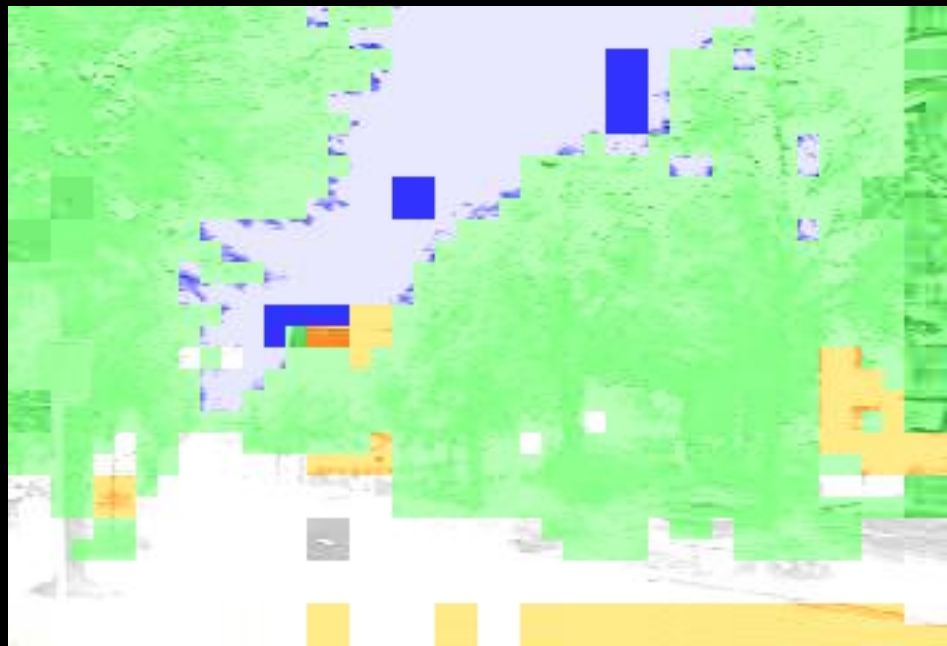
Original

43P15

High Power Blocks for Frame Difference

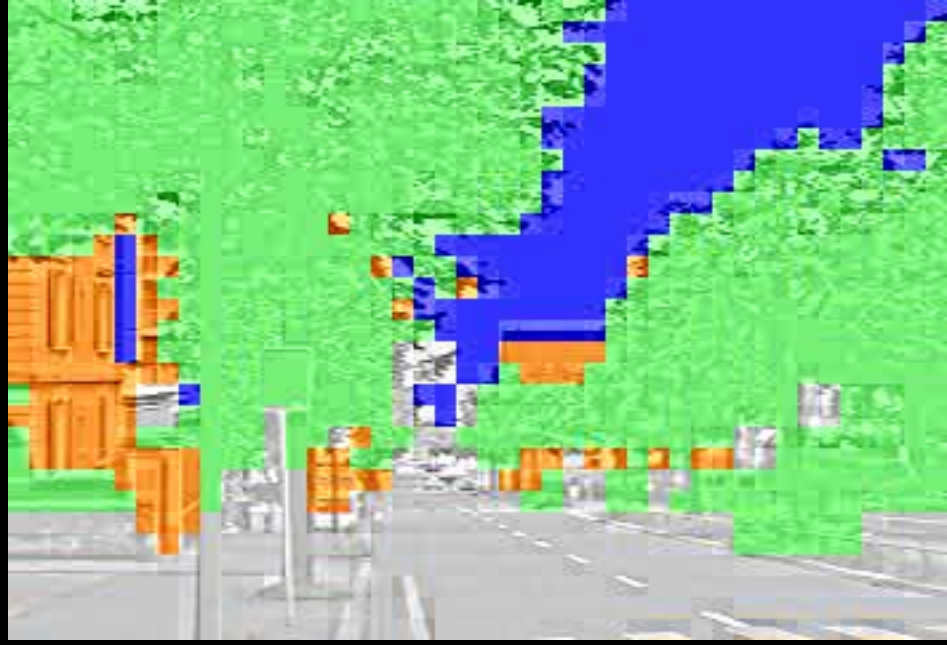
Motion Vectors and DCT coefficients

2D Recognized Results

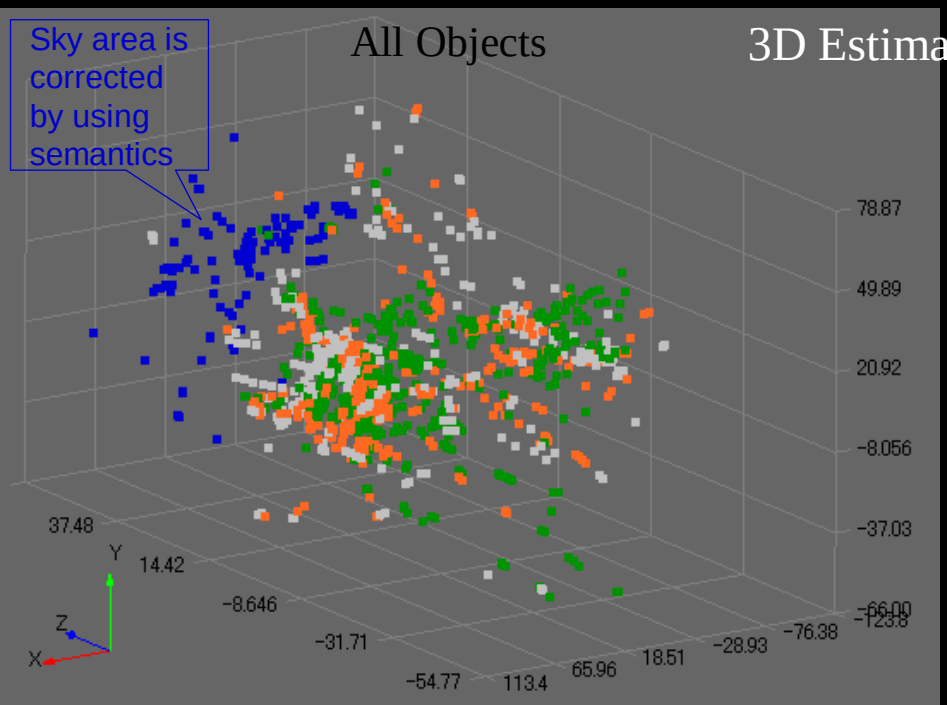


Original

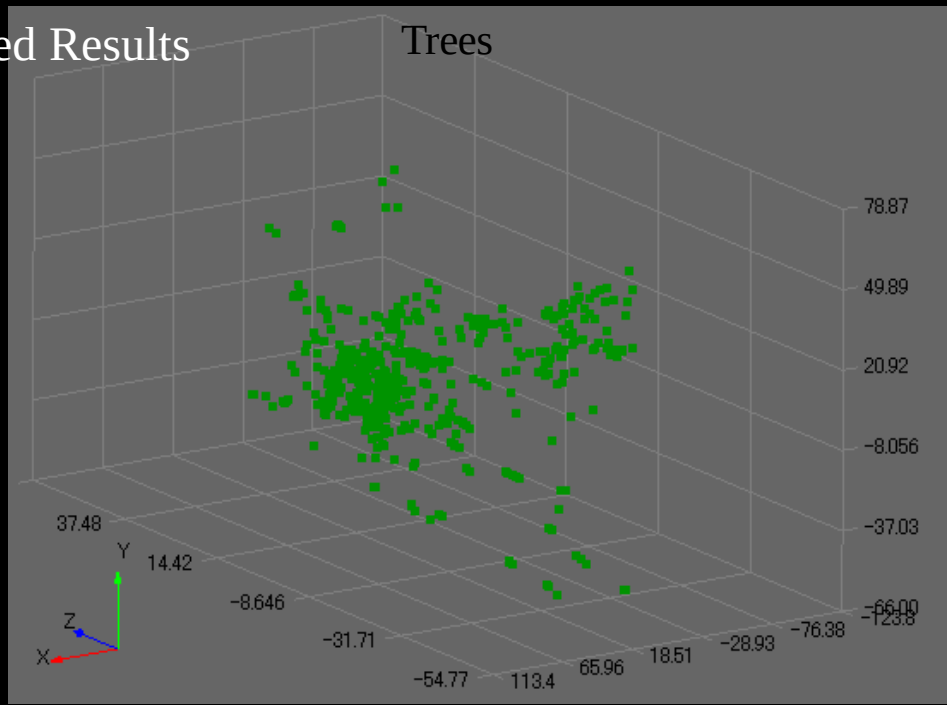
1I1



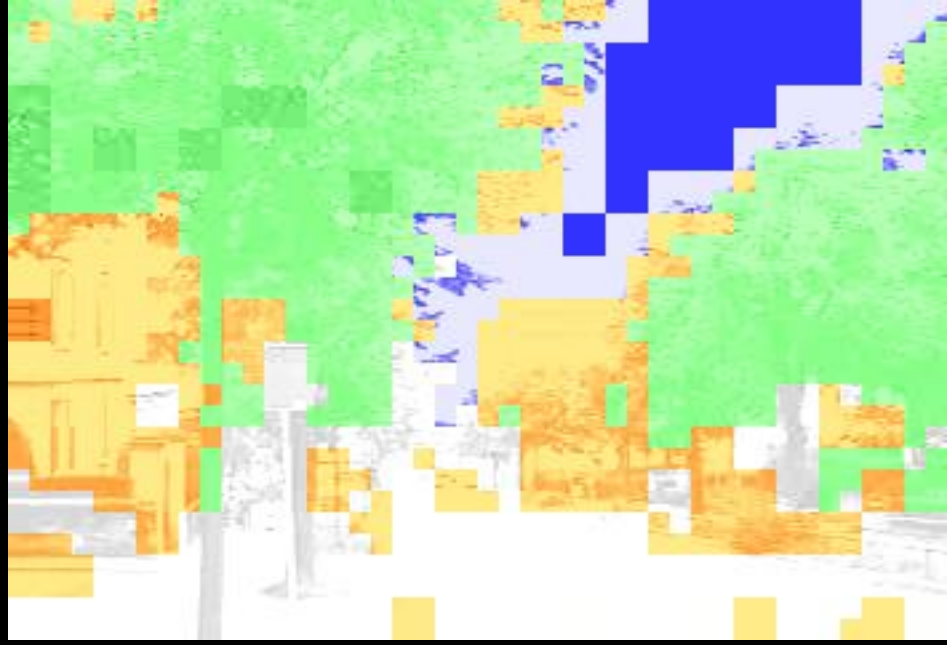
2D Recognized Results



3D Estimated Results



Original



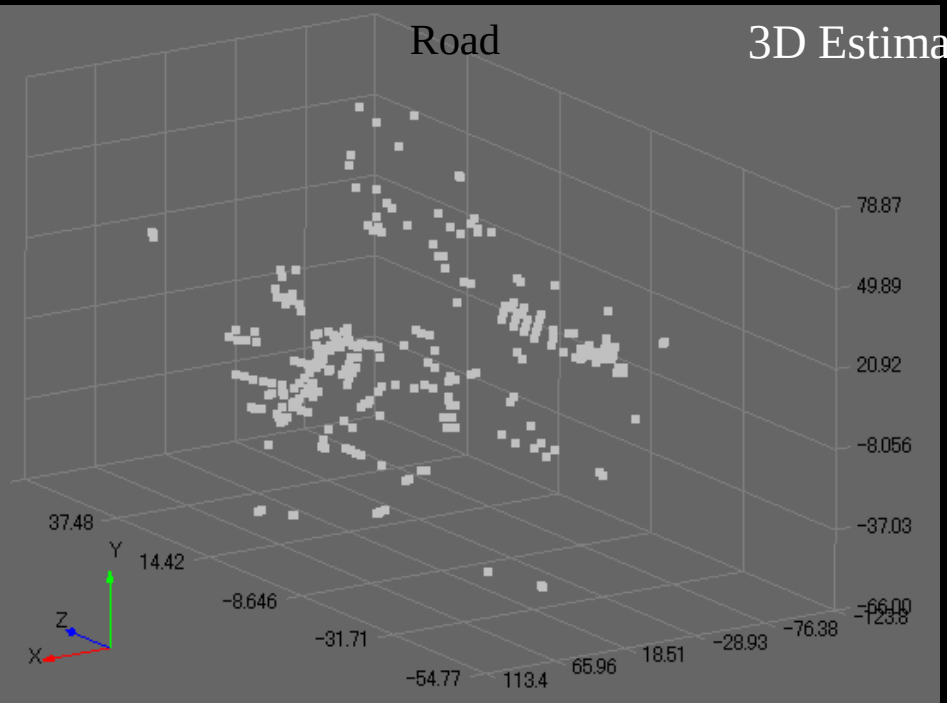
2D Recognized Results

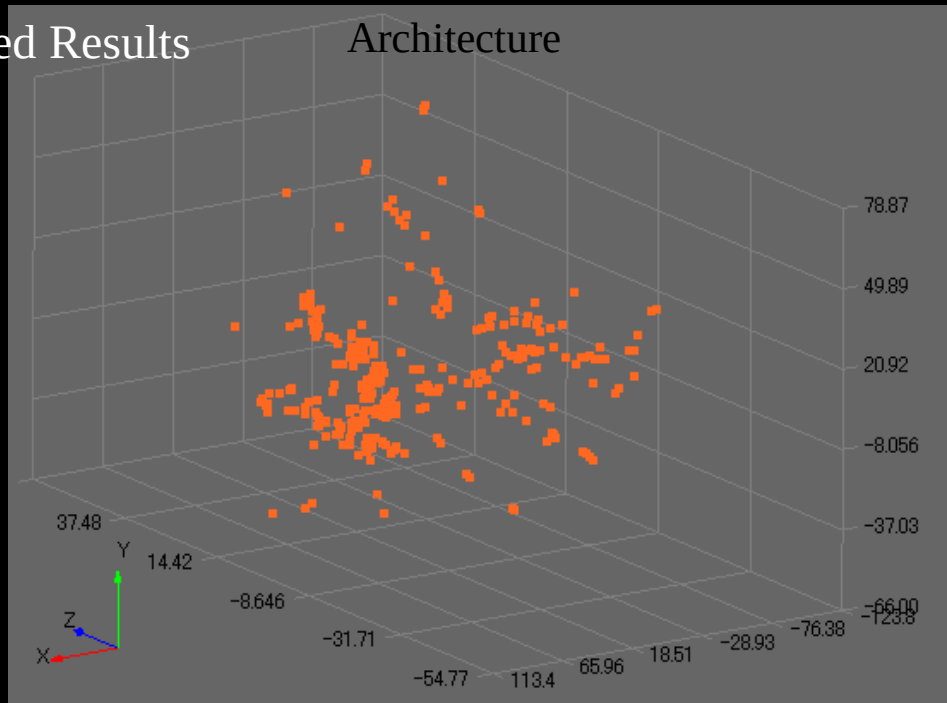
4P2

Road

3D Estimated Results

Architecture



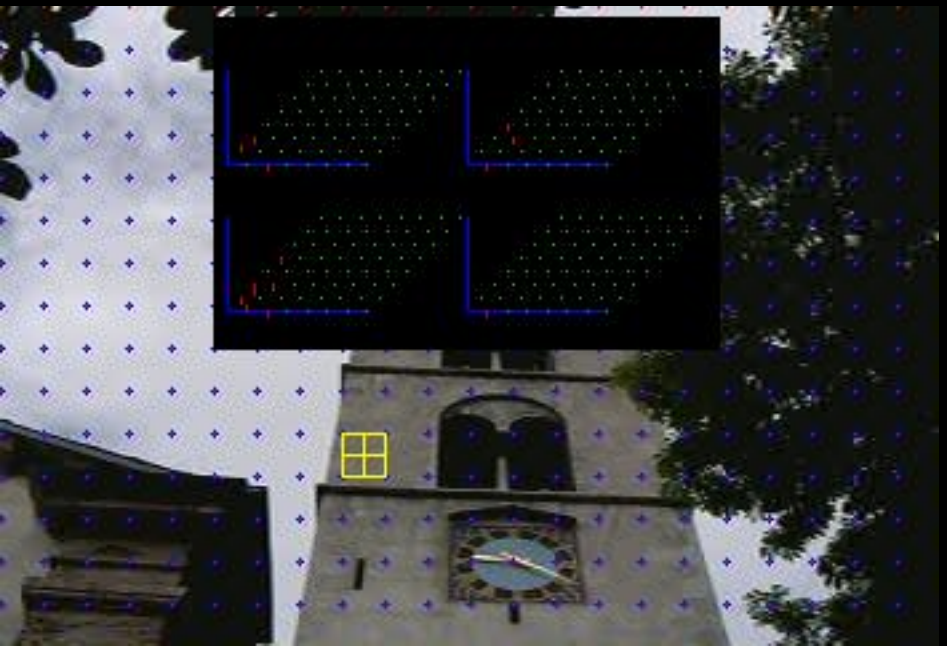


Original

High Power Blocks for Frame Difference

1I1

Motion Vectors and DCT coefficients



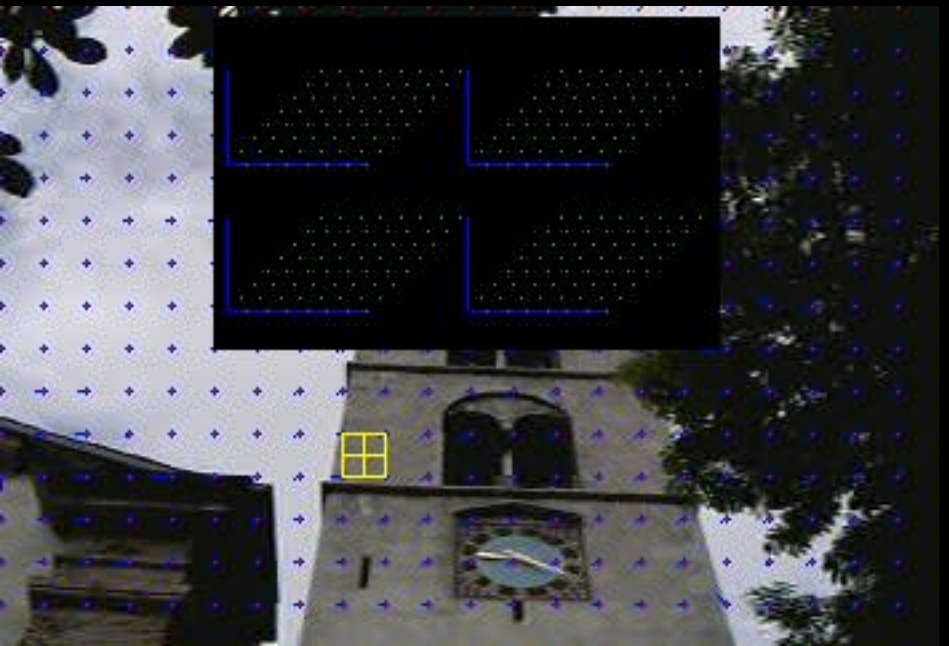
2D Recognized Results

Original

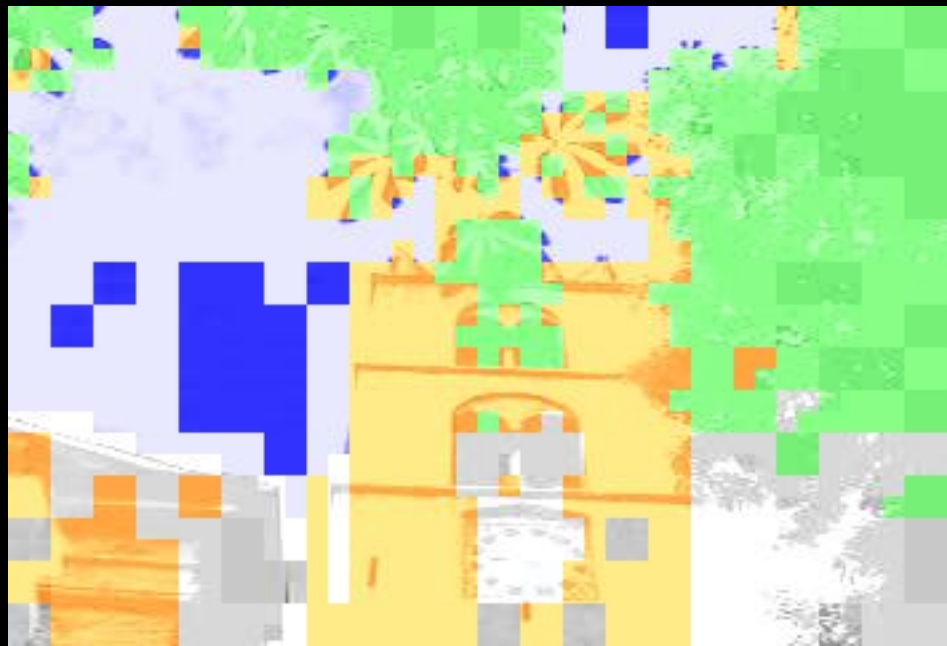
4P2

High Power Blocks for Frame Difference

Motion Vectors and DCT coefficients



2D Recognized Results

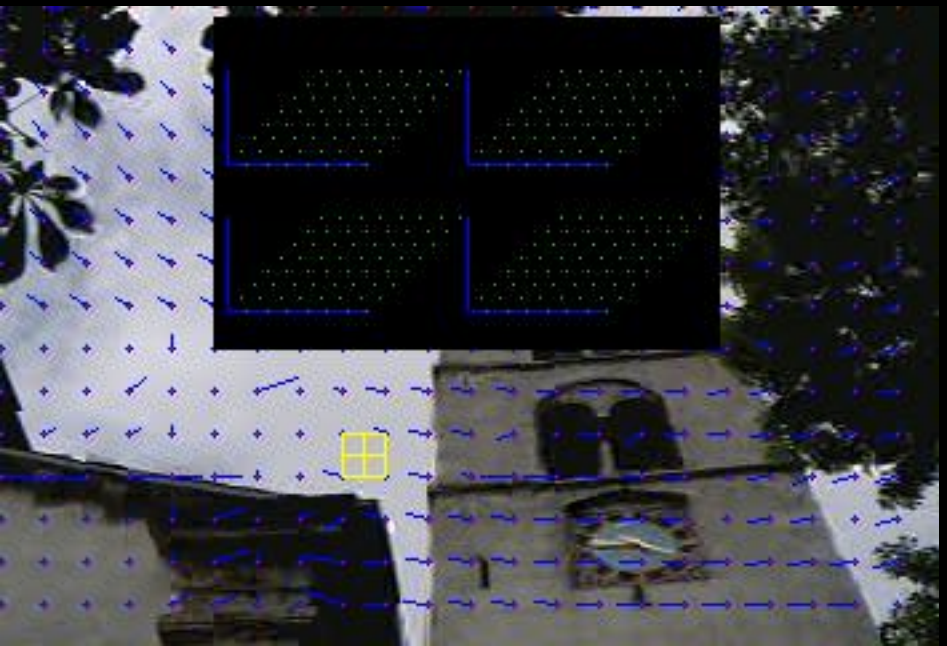


Original

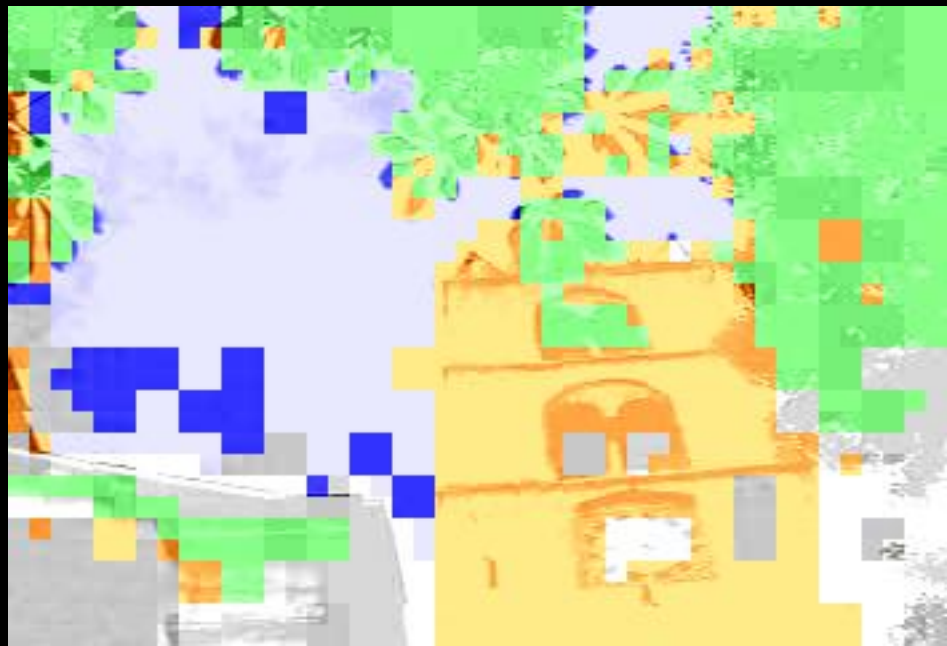
28P10

High Power Blocks for Frame Difference

Motion Vectors and DCT coefficients



2D Recognized Results

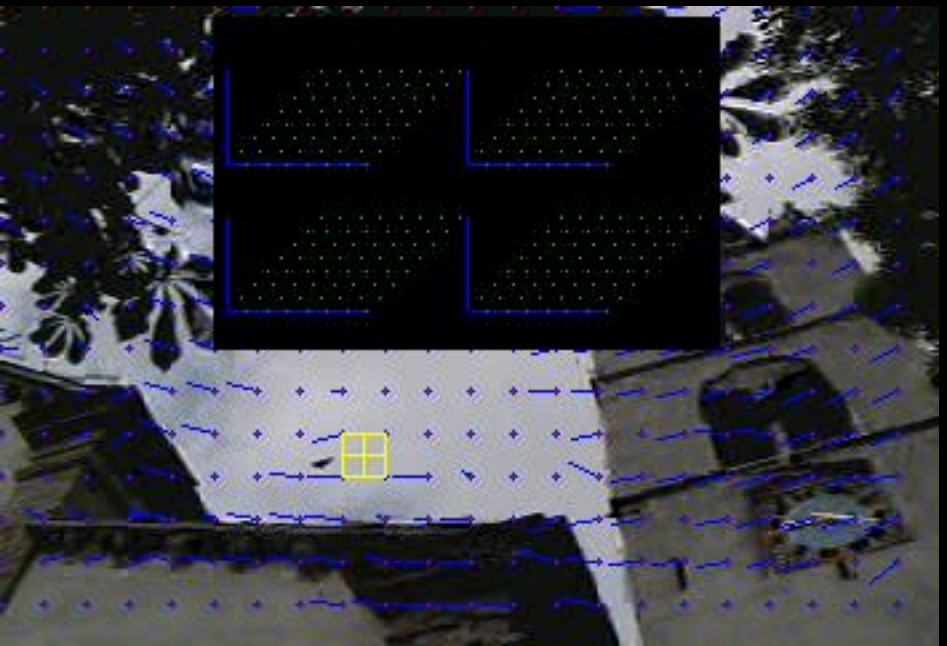


Original

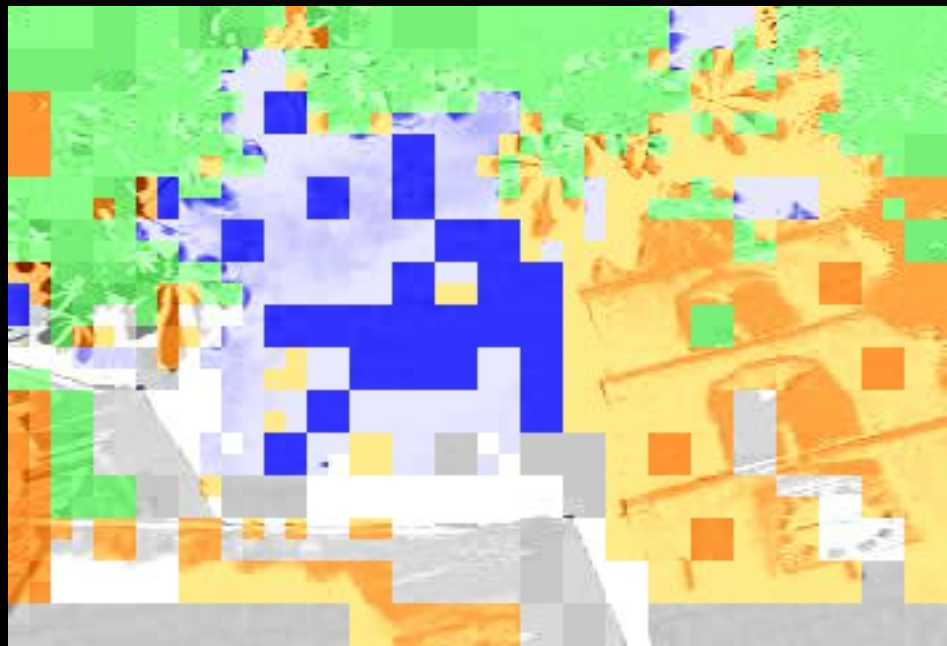
46P16

High Power Blocks for Frame Difference

Motion Vectors and DCT coefficients



2D Recognized Results



Original

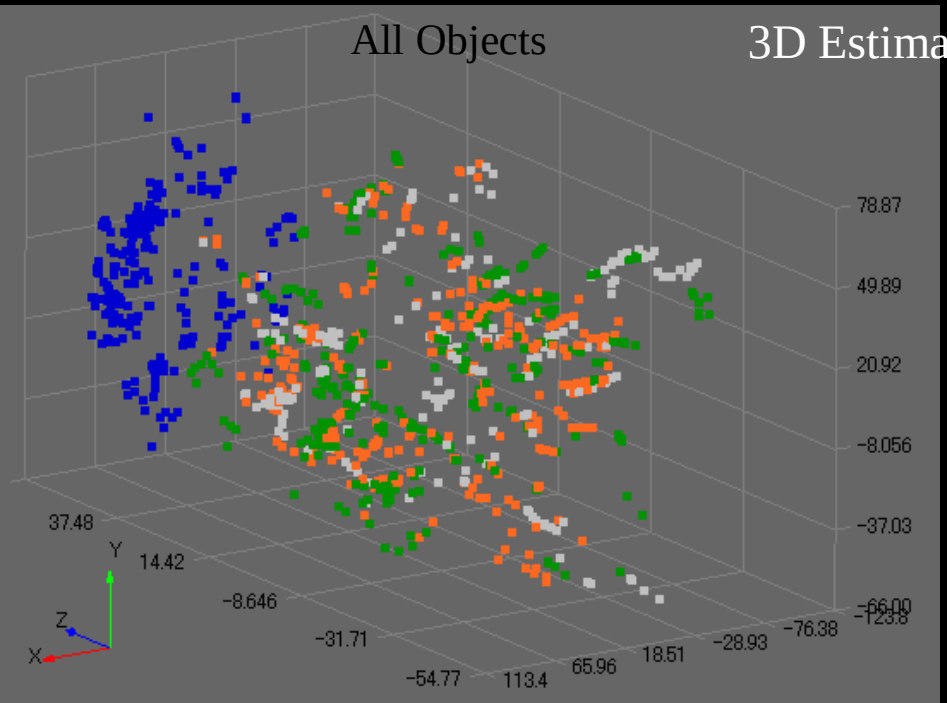
1I1

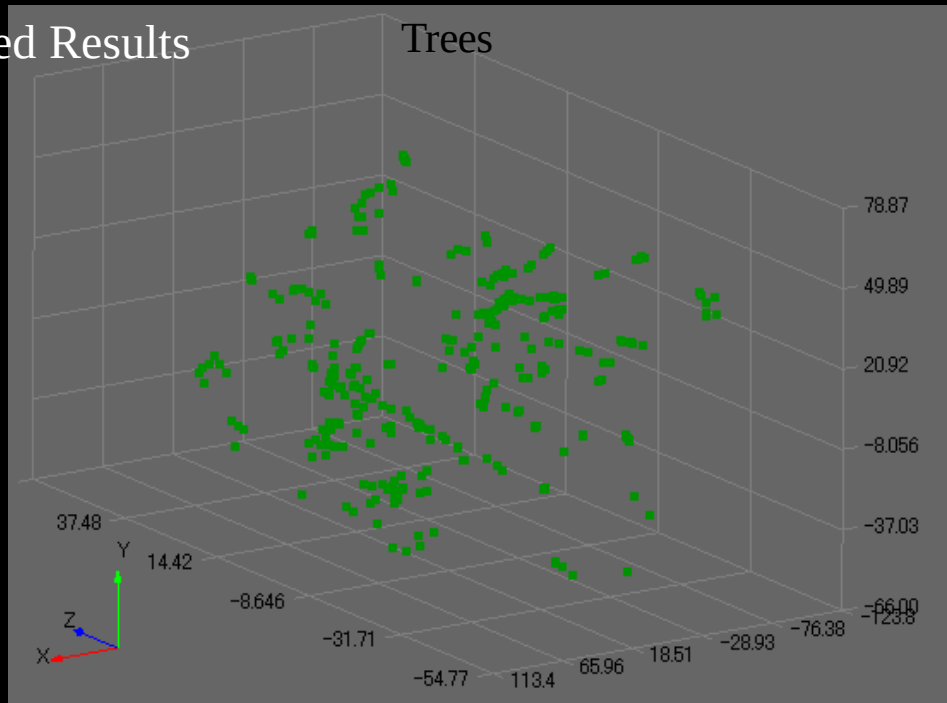
2D Recognized Results

All Objects

3D Estimated Results

Trees





Original

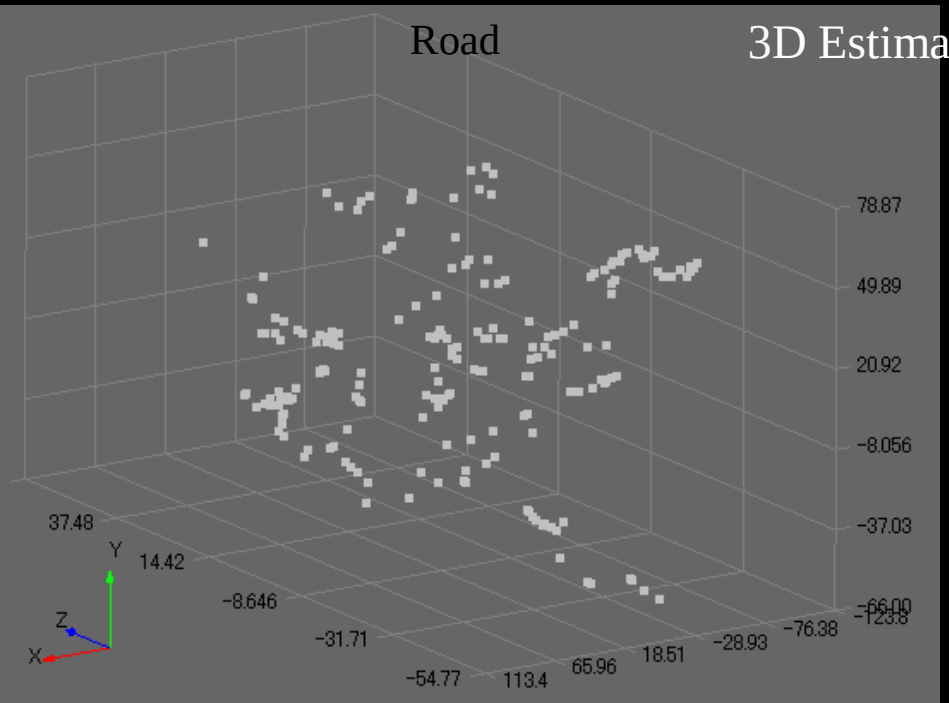
2D Recognized Results

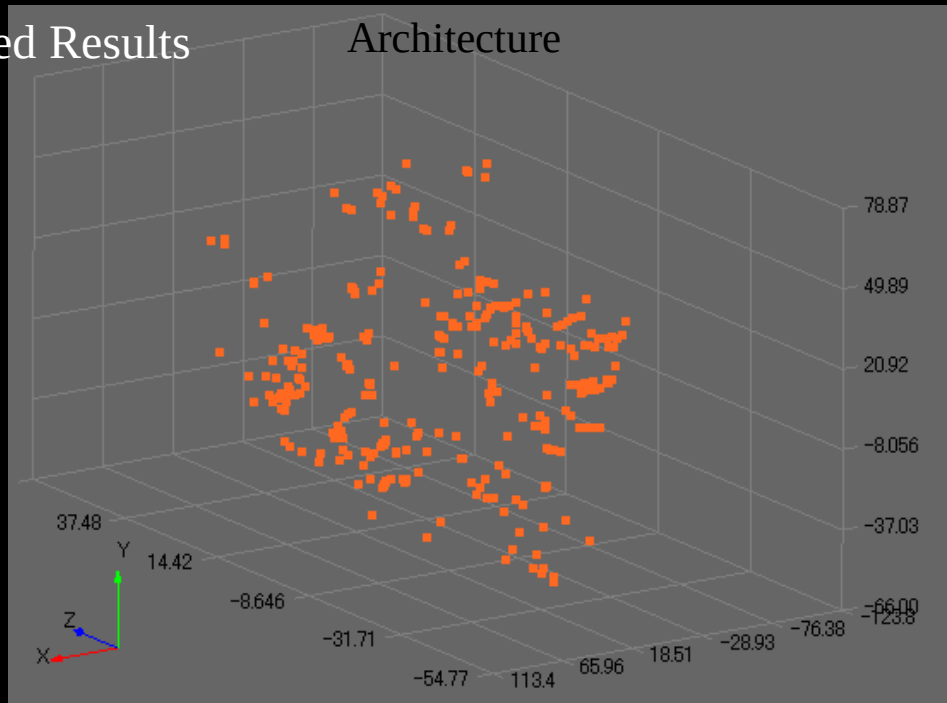
4P2

Road

3D Estimated Results

Architecture





Concluding Remarks

An Investigation on the possibilities to incorporate the factorization techniques to the MPEG-based scene class recognition was shown. There precisely exist some specific constraints for the landscape scenes especially when the real time video camera use. Negative ones are the small frame length and the computational amounts. The positive ones are as follows. As the objective of this article is to determine the scene class automatically, it is not required for the system to compute high resolution for the reconstructed 3D scene structure. Moreover, the maximum number of the scene classes defined in advance is around five.

On the other hand, there is a problem that the 3D reconstruction mechanism becomes harder when the perspective view projection is assumed. As Kanade et al. has proposed in the 90's, the factorization method seems to be very simple for the 3D reconstruction from the video images. However, for the calculation, the feature points are required to be extracted in the pixel level and the orthogonal projection was basically introduced for the simple realization. This orthogonal projection is known as the simplest class of MAP (Metric Affine Projection) . For the application to the landscapes, some new ideas would be required.

Regarding the feature points, from the aspects of MPEG Based Vision, the author has used the motion vectors extracted from block of pixels based calculation. Then the number of feature points in each frame can be set to the constant number and the 2D position of each is fixed, although the individual feature point moves from frame to frame on the specified block addresses. But this preset condition yields the rank problem because the temporal depth of measurement matrix cannot take enough values (in this paper, it is set to 20 which is equal to the number of predictive frames which hold forward motion vectors in two seconds. And the reason of the two seconds is ascribed to the practical observation time when the mechanism is installed on the handy video equipment. And another reason is that the user of the handy camera equipment usually use pan, tilt and the result of the captured scene cannot hold the same sight so long).Moreover, the number of blocks is 1320 when using MPEG-1 standard. Then totally, the measurement matrix sometimes becomes inappropriate for the SVD because of the rank problem.

In this article, the block-based factorization was proposed. The calculations themselves have been well done and very fast as compared with the pixel-wise case. The improvement of the performance is on going.

Reference

http://homepage3.nifty.com/MikioSasaki/scene/MBV3_short_120106a.pdf

http://homepage3.nifty.com/MikioSasaki/scene/ref/TMBV.pdf

http://homepage3.nifty.com/MikioSasaki/scene/MIS.pdf

http://homepage3.nifty.com/MikioSasaki/scene/090724_IEICE_SceneClass.pdf

http://homepage3.nifty.com/MikioSasaki/scene/7_1.htm