

# Improved Dense Trajectory with Cross Streams Supplemental Material

Katsunori Ohnishi  
Graduate School of  
Information  
Science and Technology  
The University of Tokyo  
ohnishi@mi.t.u-  
tokyo.ac.jp

Masatoshi Hidaka  
Graduate School of  
Information  
Science and Technology  
The University of Tokyo  
hidaka@mi.t.u-  
tokyo.ac.jp

Tatsuya Harada  
Graduate School of  
Information  
Science and Technology  
The University of Tokyo  
harada@mi.t.u-  
tokyo.ac.jp

## 1. PARAMETER ON UCF101 SPLIT1

Following previous work that encodes CNN-based local descriptors [1], we first evaluate dimension reduction. Then, we explore the number of clusters for encoding.

### 1.1 FV

We first evaluate descriptor dimensions after compression by PCA with a fixed number of gaussian mixtures  $K = 256$ . Table 1 shows that 64-D achieves the best performance on all methods. Thus, we employ 64-D for FV.

Next we evaluate number of gaussian mixtures  $K$ . Table 2 shows that  $K = 128$  achieves the best result on TDD and TDD + CPD, while  $K = 256$  performs the best on CPD. However,  $K = 128$  on CPD shows comparable performance to  $K = 256$ . Thus, we fixed  $K = 128$  both on TDD and CPD in this paper.

**Table 1: Impact of TDD and CPD dimensions after compression with fixed  $K = 256$  in FV.**

Dimensions	32-D	<b>64-D</b>	128-D	256-D
TDD	90.3%	<b>90.4%</b>	90.4%	90.2%
CPD	90.0%	<b>90.6%</b>	90.4%	90.2%
TDD+CPD	90.2%	<b>90.8%</b>	90.5%	90.5%

**Table 2: Impact of the number of gaussian mixture  $K$  with fixed PCA dimensions of 64-D in FV.**

Clusters	$K = 32$	$K = 64$	<b><math>K = 128</math></b>	$K = 256$
TDD	89.5%	90.4%	<b>90.7%</b>	90.4%
CPD	89.9%	90.2%	90.4%	90.6%
TDD+CPD	90.2%	90.6%	<b>90.8%</b>	90.8%

## 1.2 VLAD

We also evaluate dimensions and number of clusters in VLAD. Table 3 shows that 128-D achieves the best performance on CPD and TDD + CPD. Although 128-D is not the best on TDD, it achieves comparable performance. Thus, we employ 128-D for VLAD. Next we evaluate number of k-means clusters  $K$ . Table 4 shows its result. We can see the best  $K$  is 64 on CPD and TDD + CPD.  $K = 64$  also achieves almost the same result as the best one,  $K = 32$  on TDD. Thus, we fixed  $K = 64$  both on TDD and CPD in this paper.

**Table 3: Impact of the TDD and CPD dimensions after compression with fixed  $K = 256$  in VLAD.**

Dimensions	32-D	64-D	<b>128-D</b>	256-D
TDD	91.2%	91.1%	90.9%	91.1%
CPD	90.7%	91.2%	<b>91.5%</b>	91.5%
TDD+CPD	91.5%	91.2%	<b>91.5%</b>	91.4%

**Table 4: Impact of the number of k-means clusters  $K$  with fixed PCA dimensions of 128-D in VLAD.**

Clusters	$K = 32$	<b><math>K = 64</math></b>	$K = 128$	$K = 256$
TDD	<b>91.6%</b>	91.5%	91.3%	90.9%
CPD	90.3%	<b>91.6%</b>	91.3%	91.5%
TDD+CPD	91.6%	<b>92.0%</b>	91.5%	91.5%

## 2. REFERENCES

- [1] Z. Xu, Y. Yang, and A. G. Hauptmann. A discriminative CNN video representation for event detection. In *CVPR*, 2015.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*MM '16*, October 15-19, 2016, Amsterdam, Netherlands

© 2016 ACM. ISBN 978-1-4503-3603-1/16/10...\$15.00

DOI: <http://dx.doi.org/10.1145/2964284.2967222>