WP1: Video Data Analysis

Leading : UNICT
Participant: UEDIN

Fish4Knowledge Final Review Meeting - November 29, 2013 - Luxembourg

- **Fish Detection**: Background/foreground modeling algorithms able to deal with complex domains
- **Fish Tracking**: Tracking algorithms to match objects with unpredictable trajectories and in cluttered scenes
- **Fish Recognition**: Methods to recognise fish species by integrating multiple 2D perspectively distorted views over time

- Reliable background and foreground modeling for dealing with highly complex domains featured by:
  - Multimodal backgrounds and periodic movements
  - Light variability due to the light propagation in water as affected by the water surface shape
  - Low quality videos in terms of both spatial and temporal resolution
  - Atmospheric phenomena, murky water and biofouling and video compression affecting video frame quality

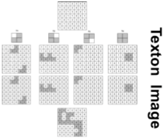
- Background modeling:
  - Using a fixed form of the *pdf* (e.g. AGMM) for background modeling shows evident limitations (**Year 1**)
  - Modeling background pixels with a set of neighbourhood samples (e.g. VIBE) instead of an explicit pixel model outperforms the above approaches (**Year 2**)
- Background movements and luminosity changes are the main causes of performance's decrease

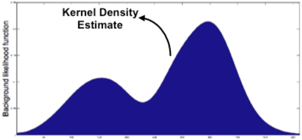
Algorithms must balance the trade-off between accuracy and efficiency

- **Description:** Data-driven Kernel Density Estimation for joint domain-range background and foreground models
- **Peculiarities:**
  - Non parametric kernel density estimator
  - Spatial Information
  - Texture Features
  - Explicit Foreground Model
- **Main limitation:**
  - Efficiency

$$\text{Each pixel } p_{(x,y)} \rightarrow v_p = (\underbrace{x, y}_{\text{Domain}}, \overbrace{s_1, s_2, s_3, e_T}^{\text{Range}})$$

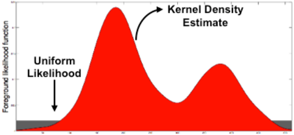
Spatial and texture information





$$P(\mathbf{p}|\psi_b) = \frac{1}{n}\sum_{i=1}^{n} \varphi(\mathbf{p} - \mathbf{b}_i)$$

$$M(x,y) = \begin{cases} 0 & \text{if } -\ln \frac{P(\mathbf{p}|\psi_b)}{P(\mathbf{p}|\psi_f)} > T \\ 1 & \text{otherwise} \end{cases}$$



$$P(\mathbf{p}|\psi_f) = \frac{1}{m}\sum_{i=1}^{m} \varphi(\mathbf{p} - \mathbf{f}_i)$$

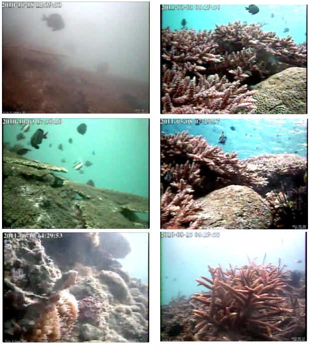
To appear on CVIU

Fish Detection
Performance Evaluation

- Datasets:
  - 17 underwater videos (spatial resolution ranging from 320 $\times$ 240 to 640 $\times$ 480)
  - I2R Dataset containing nine videos (with frames 120 $\times$160) acquired by a static camera
- Metrics:

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}$$

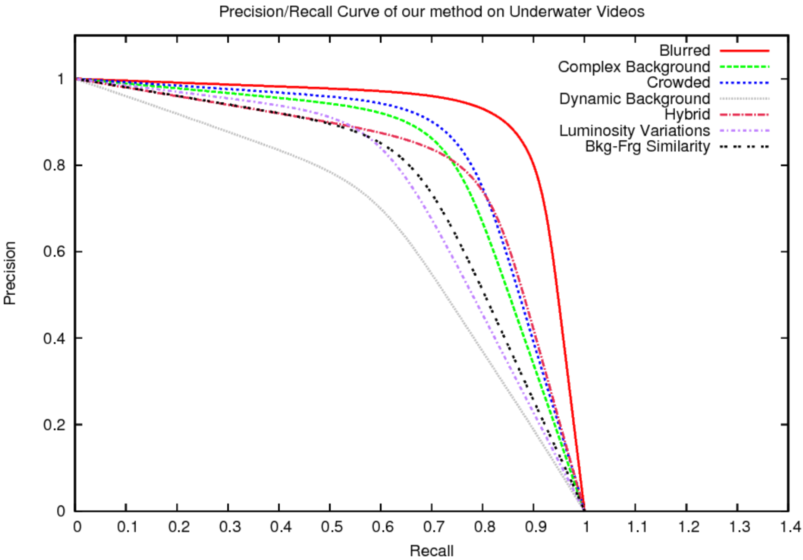
$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$



Underwater Video Dataset. From top-left to bottom-right: 1) Blurred, 2) Complex Background Texture, 3) Crowded, 4) Dynamic Background, 5) Hybrid, 6) Luminosity Change

Fish Detection

Performance of KDE on Underwater Videos



Comparison with state-of-the-art approaches on the underwater video dataset

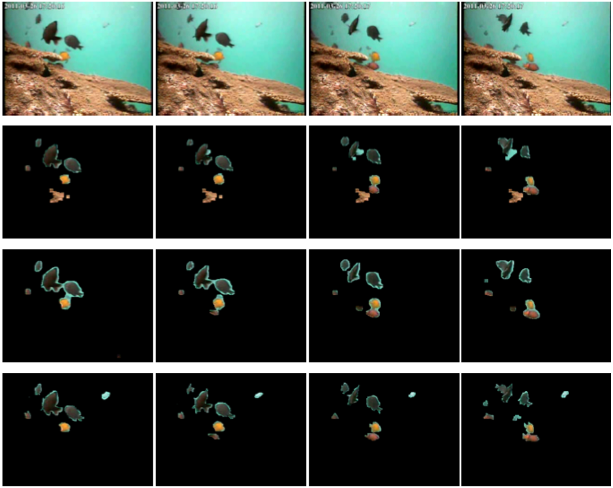
F-measures for the different background modeling approaches

Video Class	P-finder	GMM	$ZGMM$	$EIGEN$	$ML-BKG$	$KDE-RGB$	$VIBE$	*Our Method*
Blurred	78.40	79.23	81.32	80.42	73.41	90.06	86.30	92.15
Complex Background Texture	69.73	71.48	68.17	75.27	76.85	66.64	74.17	80.37
Crowded	72.85	75.32	75.56	73.65	79.83	81.72	86.83	79.84
Dynamic Background	39.92	48.23	54.48	58.99	80.60	56.78	57.98	73.41
Hybrid	64.86	65.86	66.89	76.34	77.38	78.88	73.56	84.56
Luminosity Changes	54.15	65.84	64.45	63.19	61.07	71.47	72.92	74.43
Camouflage Foreground Object	67.90	72.42	67.68	66.20	77.43	57.72	72.88	80.36
Average	63.97	68.34	68.36	70.58	75.22	71.89	74.94	80.73

Processing Times (frames/sec) on a PC powered by an Intel i7 3.4 Ghz CPU and 16GB RAM

Algorithm	320×240	640×480
P-Finder	250	60
GMM	200	50
VIBE	100	25
ZGMM	100	25
EIGEN	30	10
ML-BKG	20	3
Our recent approach	1.5	–

Fish Detection
Qualitative results



Qualitative comparison: background modeling with (from top to bottom) *VIBE* **(second row),** $ML - BKG$ **(third row) and our** *KDE* **approach (last row)**

Fish Detection
Results per zones and video classes

Image Region	P-finder	GMM	$ZGMM$	$EIGEN$	$ML-BKG$	$VIBE$
Open Sea	78.18	79.40	80.59	79.84	83.03	85.95
Corals	61.03	59.28	54.34	66.02	75.03	61.27
Rocks	64.44	73.90	68.83	64.97	76.60	77.17

F-Measure scores (in percentage) for different methods per image region

Video Class/Image Region	Open Sea	Rocks	Corals	Average
Blurred	VIBE(91.45)	–	VIBE (66.94)	**79.19**
Complex Background Texture	VIBE(79.95)	ML (86.64)	ML (87.37)	**84.65**
Crowded	VIBE (88.67)	ML (80.32)	–	**84.49**
Dynamic Background	VIBE (82.10)	ML (83.11)	ML (86.98)	**84.06**
Hybrid	EIGEN (80.16)	–	ML (77.14)	**78.65**
Luminosity Changes	VIBE (85.95)	ML (77.14)	ML (84.68)	**82.59**
Camouflage Foreground Object	VIBE (88.52)	ZGMM (86.29)	GMM (65.22)	**80.01**
Average	**85.25**	**82.70**	**78.05**	

Best performance (in terms of F-Measure) per video class and image region

Fish Detection

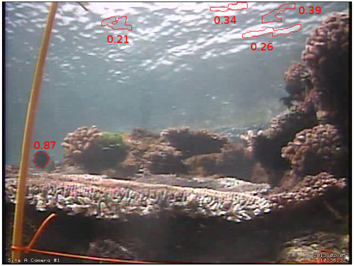
Comparison with state-of-the-art approaches on I2R dataset

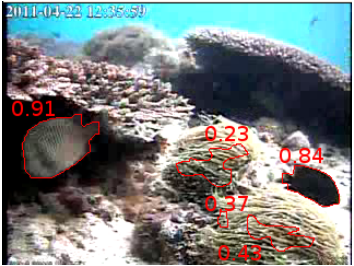
F-measures for the different background modeling approaches on the I2R Dataset

Video	KDE-RGB [2]	SILTP [28]	VKS rgb [27]	VKS Lab plus SILTP [27]	Our Method
AirportHall	61.34	68.02	70.44	71.28	69.23
Bootstrap	74.64	72.90	71.25	76.89	76.47
Curtain	97.73	92.40	94.11	94.07	94.89
Escalator	65.41	68.66	48.61	49.43	72.02
Fountain	51.32	85.04	75.84	85.97	83.21
ShoppingMall	60.36	79.65	76.48	83.03	78.54
Lobby	67.79	79.21	18.00	60.82	66.34
Trees	66.75	67.83	82.09	87.85	81.89
WaterSurface	81.57	83.15	94.83	92.61	92.51
Average	69.66	77.43	70.18	77.99	79.46

Fish Detection
Discriminate fish from background objects

- **Perceptual Organization**:
  - Gestalt Laws:
    $E\left[\partial R\right] = \frac{-\int\int_R f(x,y)dxdy}{L(\partial R)}$
- **Features**:
  - *Intraframe*: e.g. Boundary complexity, color contrast on the boundary, superpixel straddling ;
  - *Interframe*: e.g. Motion on boundary, motion homogeneity, kinematic features extracted from affine motion model, etc.
- **Performance**:
  - SVM-RBF classifier
  - Two datasets: fish and humans from I2R. About 1500 hand labeled detections.
  - Average misclassification rate (MCR) obtained with a 5-fold cross-validation: 4.34%

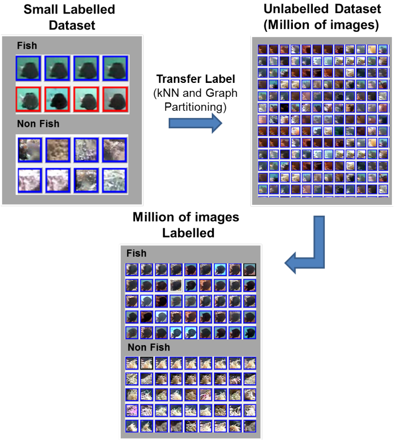




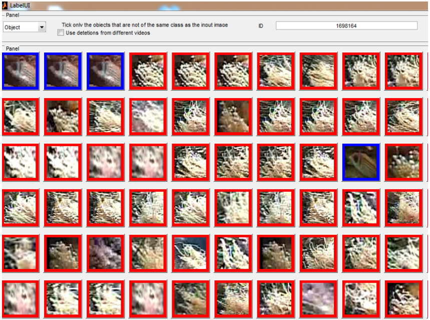
Fish Detection

Discriminate fish from background objects

A big data perspective: How to exploit the 1.4×10^9 detections to filter out bad detections?



Example



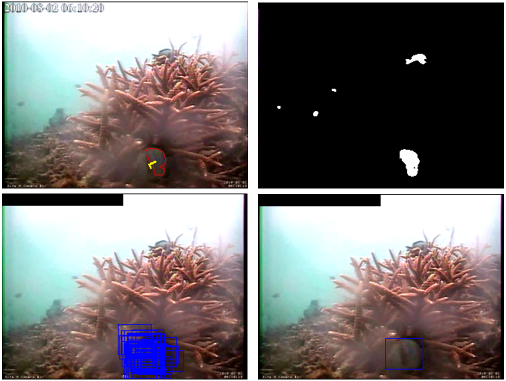
Results on 14M unlabelled images

Accuracy on test set of 10K images is ~94%

- Underwater fish tracking:
  - Fish deformations and orientation
  - Similarity between fish of same species
  - Low frame rate
- Covariance modeling
  - Spatial and statistical features
  - Position, color and gradient features

- Covariance-based tracker
  - Tracking-by-detection
  - Heuristic search area
  - Cannot fix detections
  - Occlusion: single blob
  - Faster (~0.05 s/obj.)
- Covariance particle filter
  - Weights: covariance and motion
  - Particles $\rightarrow$ search area
  - Can find object without detection
  - Can handle “touching” occlusion
  - About $10\times$ slower

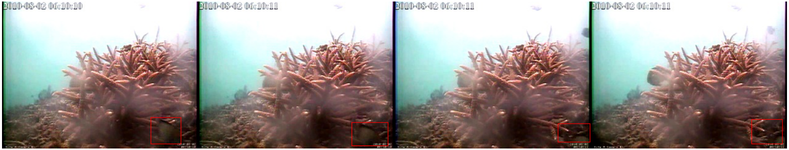
Fish Tracking
Covariance particle filter



Particle filter with covariance in action. From top-left to bottom-right: 1) Detection constrained by the background modeling, 2) Background/foreground mask, 3) Object particles (describing search area), 4) Location estimated by the particle filter.

Fish Tracking
Covariance particle filter

Covariance particle filter is able to follow object when motion detection fails...

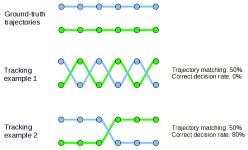


...although sometimes it follows background areas.

Fish Tracking

Performance evaluation of covariance-based tracker

- Matching Counting Rate (MCR).
- Average Trajectory Matching (ATM)
- Correct Decision Rate (CDR)



		COV			*COVPF*		
Video	**Objects**	**ATM**	**CCR**	**CDR**	**ATM**	**CCR**	**CDR**
1	1058	0.75	0.70	0.74	0.50	0.68	0.93
2	3072	0.92	0.51	0.81	0.84	0.53	0.93
3	16321	0.66	0.67	0.77	0.56	0.65	0.65
4	1927	0.73	0.56	0.80	0.69	0.55	0.89
5	1284	0.64	0.59	0.67	0.48	0.59	0.78
6	1656	0.70	0.55	0.66	0.56	0.52	0.87
7	5477	0.66	0.72	0.75	0.71	0.74	0.77
8	820	0.95	0.90	0.73	0.80	0.80	0.75
9	1447	0.88	0.66	0.73	0.84	0.63	0.83
10	1903	0.84	0.57	0.70	0.80	0.53	0.75
Avg		*0.77*	*0.64*	*0.74*	*0.68*	*0.62*	*0.82*

Comparison between original tracker and particle filter version on ground-truth videos.

Fish Tracking
Performance evaluation

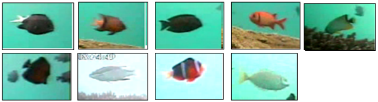
		COV			*COVPF*		
Video	**Objects**	**ATM**	**MCR**	**CDR**	**ATM**	**MCR**	**CDR**
1	344	0.86	0.85	0.84	0.86	0.85	0.88
2	260	0.84	0.85	0.83	0.85	0.85	0.88
3	121	0.75	0.71	0.81	0.80	0.76	0.83
Avg		*0.81*	*0.80*	*0.83*	*0.83*	*0.82*	*0.86*

Comparison between original tracker and particle filter version on high-res Aquacam videos.

- Evolution of fish recognition.
- Latest fish recognition component.
- Result refining after classification.

The species of this image is *Dascyllus reticulatus*.

NOT



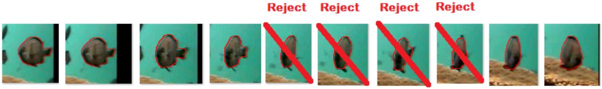
Fish Recognition

Fish recognition component Release 2 (Sep. 2012)

The species of this **trajectory** is *Dascyllus reticulatus*.

NOT

Fish Recognition
Fish recognition component Release 3 (Jul. 2013)



The species of this **trajectory** is *Dascyllus reticulatus*.

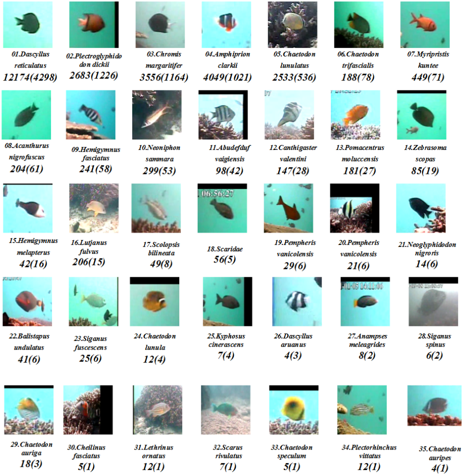
NOT

We reject some less confident recognition results.

This is a **valid** fish with the probability of 0.8907.

Fish Recognition

Fish ground-truth dataset of top 35 species



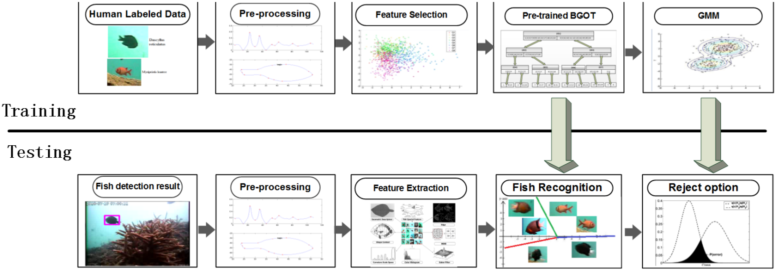
35 species

27470 detections

(8780 trajectory)

Only use top 23 species (27370 detections, 8756 trajectories).

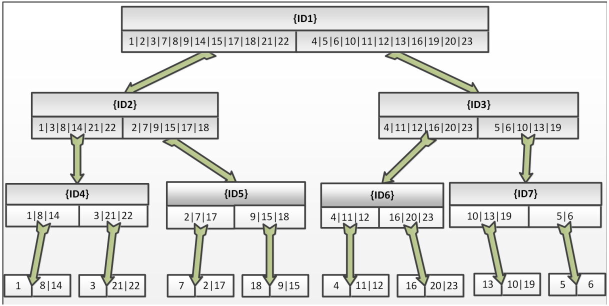
Fish Recognition
Fish recognition workflow



69 features (2626 dimensions)

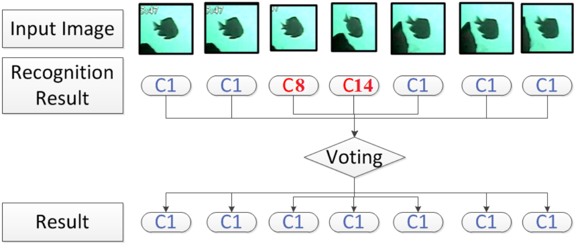
- Color
  - Normalized Red / Green histogram
  - H component histogram in HSV space
- Boundary
  - Curvature tail area ratio / Density
  - Moment Invariants / Affine Moment Invariants
  - Fourier transform
- Texture
  - Co-occurrence matrix
  - Histogram of oriented gradients
  - Gabor filter

- Balance-Guaranteed Optimized Tree (BGOT)
- Arrange more accurate classifier at a higher level.
- Keep the hierarchical tree balanced.
- Leaf node is a multi-class SVM based on 1-vs-1 strategy.



Fish Recognition

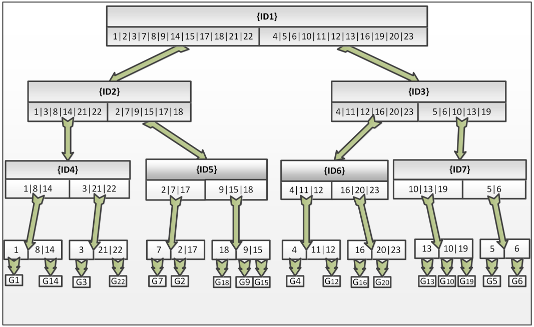
Result refining after classification: Trajectory voting



	Recall Averaged by class (%)	Precision Averaged by class (%)	Percentage of recognised fish (%)
multi-SVM	72.1	79.3	96.8
BGOT	75.3	81.9	97.0

Fish recognition result with Trajectory Voting

- Reject unlikely fish from the BGOT result.
- Tradeoff between precision and recall.
- Reduce error accumulation.



- Reject unknown species & misclassifications.
- Use specialised class model.
- Reject low probability classifications.

Algorithm	AP (%)	AR (%)
BGOT baseline	56.5 ± 2.5	$\mathbf{91.1 \pm 2.2}$
BGOT+SVM prob. rejection	59.0 ± 2.7	90.9 ± 2.3
BGOT+soft-deci. rejection	58.9 ± 2.7	90.7 ± 2.3
BGOT with GMM rejection	$\mathbf{65.0 \pm 2.7}^*$	88.3 ± 3.0

Here use 15 species as training and 8 other species (plus samples from 15 species) as testing. * means significant improvement with 95% confidence by t-test.

- Background modeling results beyond the state of the art, both in underwater videos and in standard datasets (e.g., I2R)
- Novel approach for discriminating objects of interest from the background
- A covariance particle filter able to handle multi-object occlusions and to track effectively objects with 3D complex and unpredictable trajectories
- Novel methods for recognising deforming similar shapes (fish) in 3D under variable lighting conditions

Thank you!!!

Questions?