Evaluation of Color STIPs for Human Action Recognition

Ivo Everts Jan van Gemert Theo Gevers
Intelligent Systems Lab Amsterdam, University of Amsterdam

Universiteit van Amsterdam

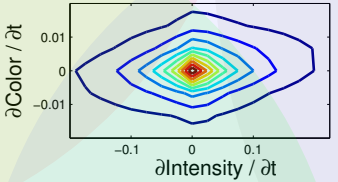
1. Why Color for Action Recognition?

Because:

Motion is the dominant cue in action recogition

Color gives:

-**More** Motion observations
-**Better** Motion estimation

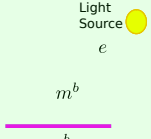


2. Reflection Model and Color Space

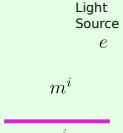
Shafer85

Di-chromatic Model, $\mathbf{f} = e(m^b\mathbf{c}^b + m^i\mathbf{c}^i)$

Diffuse Reflection



Specular (interface) Reflection



e:Energy
c^i:Body Color
c^b:Interface Color
m^i:Scalar
m^b:Scalar

Opponent color space

$$\begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} = \begin{pmatrix} R-G \\ R+G-2B \\ R+G+B \end{pmatrix}$$

	Intensity	Chromatic	Norm.-Chromatic	Hue
Representation	O_3	$[O_1, O_2]$	$\left[\frac{O_1}{O_3}, \frac{O_2}{O_3}\right]$	$\frac{O_2}{O_1}$
Invariant to	-	Highlights	Shadows	Hl. & Sh.
Reference	I	C	N	H

3. Multi-channel STIP Detectors

Laptev-IJCV05
Dollar-ICCVw05

Multi-channel Harris STIPs

$g(\cdot;\cdot,\cdot)$:3D Gauss f^j:image channel j n_c :nr channels

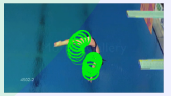
$$V = (V^1, V^2, ..., V^{n_c})^T \quad V^j = g(\cdot;\sigma_o,\tau_o) * f^j(\cdot)$$

$$V_d = (V_d^1, V_d^2, ..., V_d^{n_c})^T, d \in \{x, y, t\}$$

$$S = g(\cdot;\sigma_i,\tau_i) * \begin{pmatrix} V_x \cdot V_x & V_x \cdot V_y & V_x \cdot V_t \\ V_y \cdot V_x & V_y \cdot V_y & V_y \cdot V_t \\ V_t \cdot V_x & V_t \cdot V_y & V_t \cdot V_t \end{pmatrix}$$

Multi-channel Gabor STIPs

$$R = \sum_{j=1}^{nc} (g(\cdot;\sigma_o) * h_{ev}(\cdot;\tau_o) * V^j)^2 + (g(\cdot;\sigma_o) * h_{od}(\cdot;\tau_o) * V^j)^2$$



4. Multi-channel STIP Descriptors

KlaserBMVC08

g:3D Gradient, P: nx3 matrix of polyhedron face centers

HOG3D $\boldsymbol{q} = (q_1, ..., q_n)^T = \frac{P \cdot \boldsymbol{g}}{||\boldsymbol{g}||_2}$

Channel Integration: $\boldsymbol{g''} = \sum_{j=1}^{nc} \boldsymbol{g}^j \cdot \boldsymbol{g}^j$

Channel Concatenation: $\boldsymbol{g'} = \{\boldsymbol{g}^j\}, j = 1, ..., n_c$

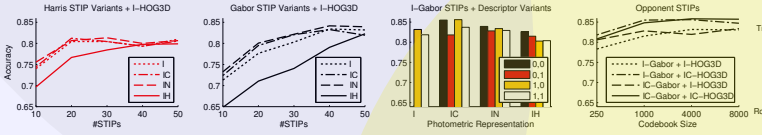
	Gradient Orientation	Gradient Direction
Channel Integration	$\mathfrak{C}_{1,1} : D/2$	$\mathfrak{C}_{1,0} : 1D$
Channel Concatenation	$\mathfrak{C}_{0,1} : n_c D/2$	$\mathfrak{C}_{0,0} : n_c D$

5. Results, UCF-Sports, UCF-11, UCF-50

Detector settings, Gabor: $\sigma_o = 2 \quad \tau_o = \sqrt{8}$ Harris: $\sigma_o = \sqrt{2}i, i \in \{2,3,4\} \quad \tau_o = \sqrt{2}j, j \in \{1,2\}$

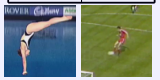
Cuboid settings, $\Delta_x = \Delta_y = 18\sigma_o, \Delta_t = 8\tau_o$ Spatio-temporal pooling: 3x3x2

Bag of features, with SVM, χ^2kernel, voc-size: 4000. Each experiment repeated 3x, typical stddev: 0.2 - 1%



		$I_{,0}$	$IC_{1,0}$	$IC_{0,0}$	$IN_{1,0}$	$IN_{0,0}$	$I_{,1}$	$IC_{1,1}$	$IC_{0,1}$	$IN_{1,1}$	$IN_{0,1}$
UCF11	$I-Gabor$	73.8%	77.5%	78.2%	76.0%	76.4%	71.6%	75.8%	74.2%	73.8%	74.6%
	$IC-Gabor$	73.8%	78.4%	78.1%	76.6%	76.3%	71.5%	75.4%	73.7%	73.9%	74.3%
	$IN-Gabor$	74.5%	77.5%	**78.6%**	76.7%	76.4%	72.4%	76.0%	74.6%	74.2%	74.0%
UCF50	$I-Gabor$	68.3%	71.7%	70.9%	71.2%	72.1%	68.8%	72.6%	69.7%	71.8%	72.0%
	$IC-Gabor$	68.5%	71.8%	70.8%	71.2%	71.9%	68.8%	72.4%	69.8%	71.5%	72.4%
	$IN-Gabor$	68.4%	71.8%	71.1%	71.0%	71.8%	68.5%	**72.9%**	69.9%	71.6%	72.5%

UCF Sports	
Intensity	Color
83%	86%



UCF 11	
Intensity	Color
74%	79%

UCF 50	
Intensity	Color
69%	73%

Ref.	Description	%
Reddy	Scene context + STIP(MBH)	76.9%
	Scene Context	47.6%
	STIP(MBH)	71.9%
Solmaz	Gist3D + STIP(HOG/HOF)	73.7%
	Gist3D	65.3%
	STIP(HOG/HOF)	54.3%
Sadanand	Action Bank	57.9%
	STIP(HOG/HOF)	47.9%
Here	Color STIP(HOG3D)	72.9%