Dynamic Texture Classification using Combined Co-Occurrence Matrices of Optical Flow

V. Andrearczyk & Paul F. Whelan

Vision Systems Group, School of Electronic Engineering, Dublin City University, Glasnevin, Dublin 9, Ireland

Abstract

This paper presents a new approach to Dynamic Texture (DT) classification based on the spatiotemporal analysis of the motion. The Grey Level Co-occurrence Matrix (GLCM) is modified to analyse the distribution of the magnitude and the orientation of the Optical Flow which describes the motion. Our method is therefore called Combined Co-occurrence Matrix of Optical Flow (CCMOF). The potential of a multiresolution analysis of the motion is revealed by experimentation. We also demonstrate the importance of the analysis of motion in the spatiotemporal domain. Finally, we demonstrate that adding a spatiotemporal motion analysis (CCMOF) to an appearance analysis (Local Binary Patterns on Three Orthogonal Planes (LBP-TOP)) significantly improves the classification results.

Keywords: Dynamic Texture, classification, Optical Flow,co-occurrence, spatiotemporal

1 Introduction

Dynamic Texture (DT) is an extension of static texture in the temporal domain, introducing temporal variations such as motion and deformation. Doretto et al. [6] describe a DT as a sequence of images of moving scenes that exhibits certain stationary properties in time. Examples of natural DTs include smoke, clouds, trees and waves. The analysis of DTs embraces several major problems including classification, segmentation, synthesis and indexing for retrieval. Those are essential for a large range of applications including surveillance, medical image analysis and remote sensing. This paper focuses on the classification of DTs. However, the developed algorithm and ideas can be used in other DT problems. The aim in DT classification is to assign an unknown sequence to a set of DT classes.

The analysis of motion for DT classification is of particular interest due to both the logic of the approach and the positive results obtained in the literature. Motion, along with shapes and color, is a key element in the analysis of a scene by the human visual system [1]. Moreover, while watching a sequence of images, the human's brain interprets the succession of still frames as a dynamic moving scene [21]. Therefore, combining a motion and an appearance analysis seems to be a natural way of dealing with dynamics. However, the literature either focuses on the spatial distribution of the Optical Flow (OF) (*motion based*) or on the pixel intensities' distribution in the spatiotemporal domain (*statistical* and *transforms*). A significant amount of information is lost by neglecting the evolution of the motion over time. In this paper, we present two new methods for DT classification which explore the potential of the analysis of the motion in the spatiotemporal domain. We use a co-occurrence matrix approach applied on the OF field in thespatiotemporal domain. We investigate the positive impact of a temporal multiresolution analysis of the motion as well as the combination of motion and appearance analysis.

2 State of the Art

The main methods developed for DT analysis can be classified in four categories, namely *statistical*, *motion based*, *model based* and *spatiotemporal transforms*.

Statistical approaches mainly use standard texture analysis in a new manner to include the temporal dimension. The Grey Level Co-occurrence Matrix (GLCM) is used by Flores et al. [8] to extract a set of features on each frames of the DT sequence. This is the most basic adaptation of a static texture analysis method to DT analysis. Hu et al. [14] also use well-known spatial descriptors on each frame but combine them with GLCM features calculated in the temporal domain to capture the correlation of neighbouring pixels in time. Boujiha et al. describe in [2] a spatiotemporal co-occurrence matrix approach, extracting co-occurrence matrices on a 3D (x, y, t) neighbourhood. In a similar manner, Zhao and Pietikäinen extend the Local Binary Pattern (LBP) to DTs by creating the LBP Volume (LBP-V) [24] and the LBP on Three Orthogonal Planes (LBP-TOP) [23]. The latter extracts LBPs on three planes: XY which is the classic spatial texture LBP as well as XT and YT which consider temporal variations of the pixel intensities.

Realising the limitations of only analysing a DT as a 3D image, **motion based** methods were developed. Extracting the motion between consecutive frames of the DT sequence can be achieved, among other methods, by Normal Flow [16], OF [4, 7, 15], or Local Motion Pattern (LMP) [9]. Various statistical analyses have been developed to extract features describing the spatial distribution of the motion field such as GLCM, Fourier spectrum, difference statistics [16], histograms [4, 9, 15] and statistics on the derivatives of the OF [7]. These motion features are often logically combined with features based on spatial statistical analyses.

Introduced by Saisan et al. in [20], **model based** methods aim at estimating the parameters of a Linear Dynamical System (LDS) using a system identification theory. This approach is designed for a synthesis problem. However, the estimated parameters can be used for a classification task [6]. Positive results were obtained in the learning and synthesis of temporal stationary sequences such as waves and clouds [6]. However, the model based approach raises several difficulties such as the distance between models lying in a non-linear space and the non-invariance to rotation and scale. Finally, it is not suitable for segmentation since it assumes stochastic and segmented DTs. Ravichandran et al. [18] overcome the view-invariance problem using a Bag of dynamical Systems (BoS) similar to a Bag of Features with LDSs as feature descriptors, obtaining precise results.

In the same manner as statistical methods, several transform approaches used in texture analysis were extended to **spatiotemporal transform** for DT analysis.Derpanis and Wildes [5] use spacetime oriented filters extracting interesting features which describe intrinsic properties of the DTs such as unstructured, static motion and transparency. In [17], Qiao and Wang use 3D Dual Tree Complex Wavelet combining an appearance with a dynamic analysis. Finally, Gonçalves et al. use spatiotemporal Gabor filters in [11].

3 Classification using Extended Plots and multiresolution

In this section, we develop a DT classification method based on the analysis of the OF. The idea of extracting the history of motion on Extended Directional and Extended Magnitude Plots developed in [16] is further explored in particular with a multiresolution analysis, different features and a rotation invariant method.

Co-occurrence matrix background The Grey Level Co-occurrence Matrix (GLCM) aims at describing the relationships between neighbouring pixel intensities by analysing their joint probability function [12, 13]. The GLCM summarizes the occurrence of pairs of pixels on a texture image. The $(i, j)^{th}$ entry of the matrix represents the number of times a pixel with intensity value i is separated from another pixel with intensity value j at a distance d in the direction θ. This matrix contains meaningful information about the distribution of the pixel intensities as indicated by the 14 Haralick's features [12, 13].

Method description

OF extraction: The OF is calculated between every two consecutive frames of the sequence using Black's algorithm [22]. As suggested by Sun et al. [16], most magnitudes being in the range zero to four pixels, the flow vectors with magnitudes greater than four pixels are regarded as noise and called *stationary points*. The other points are called *moving points*.

Quantisation: Based on [16], the magnitude and the direction are mapped into the Magnitude Plots and the Directional Plots. The magnitude is arbitrarily mapped into nine grey levels shown in Table 1. Regions of darker grey in the Magnitude Plots correspond to regions with larger normal flow magnitudes. The direction is mapped into nine grey levels as shown in Figure 1.

magnitude value	mapping
]3.555, 4]	0
]3.111 , 3.555]	31
]2.666 , 3.111]	63
]2.222 , 2.666]	95
]1.777 , 2.222]	127
]1.333 , 1.777]	159
]0.888 , 1.333]	191
]0.444 , 0.888]	223
[0 , 0.444]	255

Table 1: Magnitude mapping

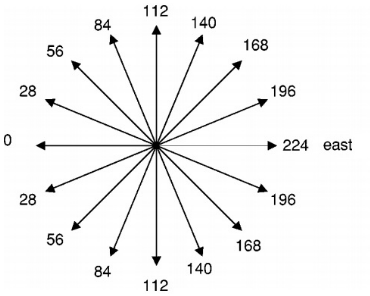


Figure 1: Quantisation of normal flow direction into 16 grey levels. (figure extracted from [16])

Extended Plots: The Extended Magnitude Plots (EMPs) and Extended Directional Plots (EDPs) [16] are calculated in order to trace the motion history. τ Magnitude Plots f_t are superimposed to construct an EMP, where $t \in [1, \tau]$ is the position of the plot in the temporal domain. The EMP $F(i, j)$ at pixel location i and j takes the value of the last moving point in $f_t(i, j)$. It takes the value 255 if $f_1(i, j), f_2(i, j), \ldots, f_\tau(i, j)$ are only stationary points. The EDP $G(i, j)$ is similarly calculated from τ Directional Plots$g_t(i, j)$. Following the setup from [16], we choose $\tau = 5$.

Multiresolution: A basic multiresolution analysis is performed, with the purpose of demonstrating the potential of this approach in a DT classification problem. The original sequences are decomposed into several temporal resolutions by applying a temporal Gaussian convolution and downsampling.

Dynamic Texture features: The co-occurrence matrices of the EMPs (9 by 9 matrices) and of the EDPs (10 by 10 matrices) are calculated and averaged over the whole sequence for each resolution. An angle $\theta = 0°$ and a distance $d = 1$ pixel between the pixel neighbours are chosen in this experiment as a proof of concept. The final features are composed of the mean co-occurrence matrices of each resolution concatenated in one single feature vector.

Rotation Invariance: In order to further estimate the robustness of the algorithm, rotation invariance is developed in the feature extraction process. The rows and columns of the co-occurrence matrices are re-organised depending on the dominant orientation of the EDPs.This is similar to rotating the OF vectors so that the dominant direction always points leftwards. Hence, similar features are extracted, for instance, from sequences of cars moving in different directions.

Results and discussions The developed algorithm is tested on the Dyntex++ database [10]. It consists of 36 classes of 100 sequences of size 50*50*50. The experimental setup is similar to [10, 18, 19]. 50 sequences are randomly selected from each class as the training set, and the other 50 sequences are used in testing. This process is repeated 100 times to obtain the average classification rate. We use a linear Support Vector Machine (SVM) classifier [3]. The classification results are summarised in Table 2 and compared to the state of the art in Table 3. The state of the art method which obtains the best results ([19]) uses PCA-cLBP (PCA on the concatenated LBP), PI-LBP (Patch-Independent), PD-LBP (Patch-Dependent) and a super histogram averaging the histograms across the sequence. Finally it uses a RBF SVM classifier. Our experiment does not aim to maximise the classification results, but rather to demonstrate the importance of the OF, the multiresolution analysis and the rotation invariance in a DT classification framework. It performs only 2.7% worse than the state of the art on the Dyntex++ database. The multiresolution analysis greatly improves the performance of our method from 74.8% to 89.7% with respectively one and four resolutions. Moreover, one could expect the classification success rate to dramatically drop with the rotation invariance since crucial information about the direction is lost. However, it is interesting to point out that we only measure a drop of 2.3% and 1.6% with respectively four and one resolutions.

number of resolutions	1	2	3	4
normal	74.8%	84.4%	87.4%	89.7%
rotation invariant	73.2%	82.5%	85.6%	87.4%

Table 2: Global classification results of our method on Dyntex ++ for different numbers of resolutions

Method	Classification rate
DL-PEGASOS [10]	63.7%
Extended Plots - 4 resolutions (our method)	89.7%
PCA-cLBP/PI-LBP/PD-LBP+super histogram + RBF SVM [19]	92.4%

Table 3: Classification results comparison with the state of the art on Dyntex++

Finally, it should be noted that some sequences are misclassified into classes with very similar motion. For instance, eight sequences of the class 27 "leaves on branches swaying with wind" are misclassified into the class 33 ("branches swaying in wind (no leaves)". Adding a spatial analysis to this method would overcome this issue.

4 Combined Co-occurrence Matrix of Optical Flow (CCMOF)

Introduction As shown in section 3, the co-occurrence matrix approach developed in [12, 13] can be used to characterise the distribution of the OF. However, creating a co-occurrence matrix from a vector image is not as straightforward as the classic GLCM extracted from pixel intensities. Indeed, the occurrence of two values must be taken into account such as the x and y components of the flow vectors or their magnitudes and directions. This is why the EDPs and EMPs were used in [16]. However, analysing the two components separately gives rise to a loss of information carried by the joint distribution. Furthermore, the quantisation of the magnitude is not as simple as the quantisation of a grey-scale image or an orientation image. Grey-scale images and orientation images are respectively defined in the bounded intervals [0,255] and [-180,180], whereas the magnitude is only limited by the size of the image. In [16] and section 3, flow vectors with a magnitude larger than four pixels were considered as noise which is not the case in many DT sequences. Extending the quantisation to the maximum magnitude that a correct flow vector can have (maximum magnitude of all the sequences, up to 25 pixels length) is not a convenient solution either. In order to overcome these issues, a new framework combining the magnitude and the orientation of the flow is developed using individual quantisation levels of the magnitudes for each sequence.

Method description The framework of our CCMOF method is illustrated in Figure 2. The OF is calculated between every two consecutive frames of the DT sequence, using the algorithm from [22]. A sequence of 100 frames thus results in 99 motion vector images. In the second block, the magnitude and the orientation of the OF are calculated from the x and y values of the motion vectors. Subsequently, the magnitude and the orientation are jointly quantised. As shown in Figure 3, eight bins of magnitude and eight bins of orientation are used, resulting in 64 bins spanning the OF vector values. For each sequence, the magnitude is linearly quantized in eight bins in the ranges $[0, M_s]$ $([0, \frac{1}{8}M_s], [\frac{1}{8}M_s, \frac{2}{8}M_s], \ldots, [\frac{7}{8}M_s, M_s])$, where M_s is the largest magnitude

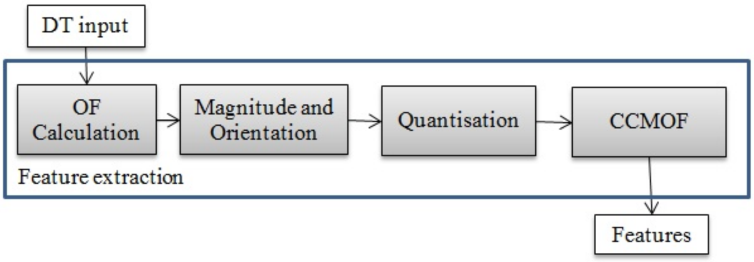


Figure 2: Feature extraction diagram of the CCMOF

of the sequence after removing large outliers. Finally, the co-occurrence of the neighbouring flow vectors is calculated, resulting in two 64 by 64 CCMOFs; one for the spatial, the other one for the temporal domain. The matrix extracted in the space domain summarises the occurrence of the pairs of neighbours on the x and y axes. This co-occurrence is calculated on every frame, then added up on the entire sequence and normalised. In the time domain, the same process is applied with neighbours on the temporal axis. These co-occurrence matrices differ from the classic GLCM as they combine two dimensions. The definition of new features is necessary in order to extract the distributions of the magnitude and of the orientation as well as their joint distribution. In total, 19 features are calculated based on Haralick's features [13]; 11 in the spatial domain and eight in the temporal domain: The *Energy*, *Contrast*, *Orientation Contrast*, *Magnitude Contrast*, *Correlation*, *Sum of Squares Variances*, *Homogeneity* and *Entropy* are extracted from both the spatial and temporal CCMOFs. The *Dominant Orientation*, *Dominant Magnitude* and *Weighted Dominant Orientation* are calculated only in the spatial domain. We need to slightly modify the calculation of the Haralick's features [13] regarding the difference n between neighbours. In Haralick's features, the difference between neighbours n is defined as Equation (1) and represents the variation in the neighbours' intensities.

$$n = |i - j|, \quad n \in \{q_1, q_2, \ldots, q_N\} \tag{1}$$

Where $i, j \in \{q_1, q_2, \ldots, q_N\}$ represent the quantised values of the pair of neighbours and $q_1, q_2, \ldots, q_N$ are the N discrete grey levels of quantization. In our case, n is chosen as the Manhattan distance between neighbour vectors (Equation (2)).

$$n = |\hat{M}_i - \hat{M}_j| + |\hat{O}_i - \hat{O}_j| \mod 4, \quad n \in \{0; 1; 2; \ldots; 11\} \tag{2}$$

Where $\hat{M}_i, \hat{M}_j$ are the quantised magnitudes of the neighbour vectors and $\hat{O}_i, \hat{O}_j$ their orientations. On Figure 3, the distance between the two vectors is $n = 2$. Finally, it was observed by examining the CCMOFs and the OF that the smallest magnitudes in a vector image are often due to noise on a static part of the DT. For instance, in a traffic DT, OF vectors are calculated on the supposedly static road with magnitudes close to zero. Their orientation is therefore not relevant and reduces the discriminative power of the features calculated from the CCMOFs. Therefore, the same 19 features are calculated from the CCMOFs in which the bin corresponding to the smallest magnitudes is removed (resulting in 7 by 8 matrices). A total of 38 features is thus extracted from each DT sequence.

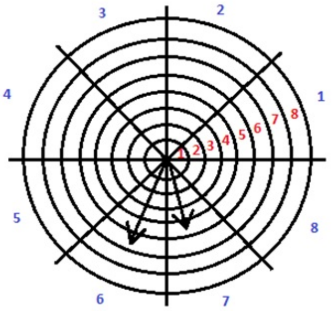
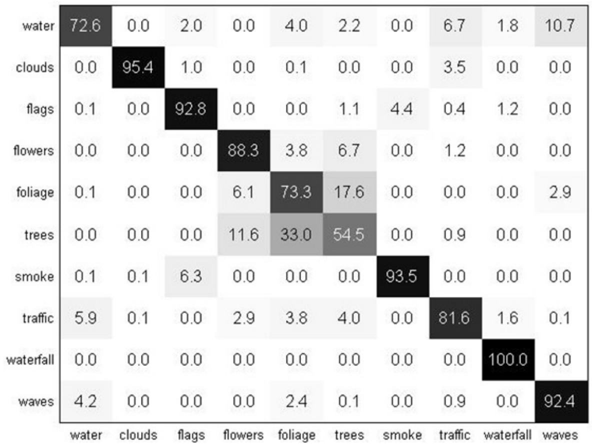


Figure 3: OF quantisation with two vector examples ($\hat{M} = 7$, $\hat{O} = 6$ (left); $\hat{M} = 6$, $\hat{O} = 7$ (right)).

DT recognition, as proposed in [19], should largely rely on appearance analysis as it contains the most discriminative information. Therefore, we combine our motion features with LBP-TOP features [23]. Thus, we cover the analysis of the motion (CCMOF) in the spatiotemporal domain as well as the pixel intensities' distribution (LBP-TOP) in the spatiotemporal domain.

Results and discussion A sequential feature selection is applied in order to determine those which discriminate the best between classes and to remove irrelevant and redundant features.We use a K-Nearest Neighbors (K-NN) classifier with $K = 1$. Angles $\theta = 0°$, $\theta = 90°$ (x and y axes) and a distance $d = 1$ pixel between neighbours are chosen for the construction of the co-occurrence matrices. For the testing of this method, a dataset is created using sequences from the Dyntex database. The current benchmarks (Dyntex, Dyntex++, UCLA) suffer, in our point of view, several drawbacks. We sought to rectify those by focusing on videos which exhibit only one DT and contain the same dynamic on the entire space-time domain. We also select sequences which depict different scenes, from various viewpoints. The resulting dataset contains 10 classes of DTs, each with 20 sequences of size 100*100*50. The experimental setup is the same as in section 3. 10 sequences are randomly selected from each class for the training set. The other 10 sequences are used for testing. The confusion matrix presenting the classification results using the CCMOF features in combination with LBP-TOP is shown in Figure 4 and the overall classification results are presented in Table 4. One can notice that the main misclassifications occur for very similar classes ("water"/"waves" and "foliage"/"trees") which share similar dynamics and spatial texture.



	water	clouds	flags	flowers	foliage	trees	smoke	traffic	waterfall	waves
water	72.6	0.0	2.0	0.0	4.0	2.2	0.0	6.7	1.8	10.7
clouds	0.0	95.4	1.0	0.0	0.1	0.0	0.0	3.5	0.0	0.0
flags	0.1	0.0	92.8	0.0	0.0	1.1	4.4	0.4	1.2	0.0
flowers	0.0	0.0	0.0	88.3	3.8	6.7	0.0	1.2	0.0	0.0
foliage	0.1	0.0	0.0	6.1	73.3	17.6	0.0	0.0	0.0	2.9
trees	0.0	0.0	0.0	11.6	33.0	54.5	0.0	0.9	0.0	0.0
smoke	0.1	0.1	6.3	0.0	0.0	0.0	93.5	0.0	0.0	0.0
traffic	5.9	0.1	0.0	2.9	3.8	4.0	0.0	81.6	1.6	0.1
waterfall	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
waves	4.2	0.0	0.0	0.0	2.4	0.1	0.0	0.9	0.0	92.4

Figure 4: Confusion matrix of CCMOF + LBP-TOP on the developed database

Method	Classification rate
CCMOF	68.0%
LBP-TOP [23]	81.7%
LBP-TOP + CCMOF	84.4%

Table 4: Classification results of CCMOF and LBP-TOP on the developed database

The motion analysis using the CCMOF approach significantly improves the classification of our dataset when combined with the analysis of the spatiotemporal distribution of the pixel intensities (LBP-TOP). Solely analysing the motion is not sufficient to obtain a classification as accurate as the literature. It confirms that a DT recognition system mostly relies on the spatial texture analysis [19] and the motion analysis provides complementary information that can improve the performance.

5 Conclusion

In this paper, we investigated the analysis of the OF in the spatiotemporal domain and the combination with complementary spatiotemporal features based on pixel intensities' distribution. Section 3 showed a simple use of the OF for DT classification with co-occurrence matrices as well as the potential of a temporal multiresolution analysis of the motion. Section 4 presented good results obtained with our new method CCMOF. In particular it showed the importance of the analysis of the motion in the spatiotemporal domain, the combination of an appearance and motion analysis as well as the joint analysis of the magnitude and orientation of the motion. Our CCMOF approach greatly improves the state of the art on the developed database when combined with appearance analysis (LBP-TOP).

References

[1] E. H. Adelson and J. R. Bergen. Spatiotemporal energy models for the perception of motion. *JOSA A*, 2:284–299, 1985.

[2] T. Boujiha, J.-G. Postaire, A. Sbihi, and A. Mouradi. New approach for dynamic textures discrimination. pages 1–4, 2010.

[3] C.-C. Chang and C.-J. Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2:27, 2011.

[4] J. Chen, G. Zhao, M. Salo, E. Rahtu, and M. Pietikainen. Automatic dynamic texture segmentation using local descriptors and optical flow. *Image Processing, IEEE Transactions on*, 22:326–339, 2013.

[5] K. G. Derpanis and R. P. Wildes. Dynamic texture recognition based on distributions of spacetime oriented structure. pages 191–198, 2010.

[6] G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto. Dynamic textures. *International Journal of Computer Vision*, 51:91–109, 2003.

[7] S. Fazekas and D. Chetverikov. Dynamic texture recognition using optical flow features and temporal periodicity. pages 25–32, 2007.

[8] A. B. Flores, L. A. Robles, R. M. M. Tepalt, and J. D. C. Aragon. Identifying precursory cancer lesions using temporal texture analysis. pages 34–39, 2005.

[9] P. Gao and C. L. Xu. Extended statistical landscape features for dynamic texture recognition. volume 4, pages 548–551, 2008.

[10] B. Ghanem and N. Ahuja. Maximum margin distance learning for dynamic texture recognition. pages 223–236. 2010.

[11] W. N. Gonçalves, B. B. Machado, and O. M. Bruno. Spatiotemporal gabor filters: a new method for dynamic texture recognition. *arXiv preprint arXiv:1201.3612*, 2012.

[12] R. M. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67:786–804, 1979.

[13] R. M. Haralick, K. Shanmugam, and I. H. Dinstein. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, pages 610–621, 1973.

[14] Y. Hu, J. Carmona, and R. F. Murphy. Application of temporal texture features to automated analysis of protein subcellular locations in time series fluorescence microscope images. pages 1028–1031, 2006.

[15] Z. Lu, W. Xie, J. Pei, and J. Huang. Dynamic texture recognition by spatio-temporal multiresolution histograms. volume 2, pages 241–246, 2005.

[16] C.-H. Peh and L.-F. Cheong. Synergizing spatial and temporal texture. *Image Processing, IEEE Transactions on*, 11:1179–1191, 2002.

[17] Y.-l. Qiao and F.-s. Wang. Dynamic texture classification based on dual-tree complex wavelet transform. pages 823–826, 2011.

[18] A. Ravichandran, R. Chaudhry, and R. Vidal. View-invariant dynamic texture recognition using a bag of dynamical systems. pages 1651–1657, 2009.

[19] J. Ren, X. Jiang, and J. Yuan. Dynamic texture recognition using enhanced lbp features. pages 2400–2404, 2013.

[20] P. Saisan, G. Doretto, Y. N. Wu, and S. Soatto. Dynamic texture recognition. volume 2, pages II–58, 2001.

[21] L. C. Sincich and J. C. Horton. The circuitry of v1 and v2: integration of color, form, and motion. *Annu. Rev. Neurosci.*, 28:303–326, 2005.

[22] D. Sun, S. Roth, and M. J. Black. Secrets of optical flow estimation and their principles. pages 2432–2439, 2010.

[23] G. Zhao and M. Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29:915–928, 2007.

[24] G. Zhao and M. Pietikäinen. Dynamic texture recognition using volume local binary patterns. pages 165–177. 2007.